

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 11, 2004, 15:58:00 ; Search time 61 Seconds

(without alignments)  
7605.865 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923  
Sequence: 1 MASPATGVLPFPPLPPLLL.....AGTVDDSSGSEPLFPPLH 2923

Scoring table: Oligo  
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2923	100.0	2923	22 AAU07054 Human Flamingo pro
2	2923	100.0	2923	23 AAM50866 Cadherin EGF LAG s
3	2923	100.0	2923	24 ABP82018 Human GPCR CELSR2
4	2894	99.0	2894	24 ABT01556 Human MDDT polypep
5	2837	97.1	2956	22 AAU07053 Human Flamingo pol
6	1958	67.0	2936	22 AAU74826 Human REPT9 prot
7	1917	65.6	2560	22 ABB11404 Human FLAMINGO 1 h
8	1623	55.5	2405	21 AAB42192 Human ORFX ORF1956
9	534	18.3	568	19 AAW51244 Human Calcitonin r

10	485	16.6	565	24 ABJ37074 Human breast cance
11	281	9.6	568	21 ABB89665 Human polypeptide
12	281	9.6	717	21 AAB56721 Human prostate can
13	220	7.5	470	24 ABP75877 Human secretory po
14	118	4.0	219	22 ABB03070 Human expressed po
15	118	4.0	219	22 AAU19681 Human novel extrac
16	118	4.0	219	22 ABB47901 Human polypeptide
17	100	3.4	111	22 ABB17070 Human nervous syst
18	100	3.4	111	22 AAU19829 Human novel extrac
19	100	3.4	111	22 ABB48049 Human polypeptide
20	82	2.8	96	22 ABB10351 Human CDNA SBO ID
21	82	2.8	96	22 AAU18132 Human novel uterin
22	82	2.8	96	22 AAU17045 Human novel secret
23	82	2.8	96	22 AAU19897 Human novel calcin
24	82	2.8	96	22 ABB66938 Human polypeptide
25	82	2.8	96	22 ABB05759 Novel human protei
26	42	1.4	141	22 AAM89890 Human immune/haema
27	42	1.4	141	22 AAU19685 Human novel extrac
28	42	1.4	141	22 ABB47905 Human polypeptide
29	23	0.8	2707	18 AAM27161 Mouse receptor ME2
30	21	0.7	44	17 AAR87108 Protocadherin clon
31	21	0.7	45	15 AAR58866 Rat-316 cadherin p
32	20	0.7	20	23 AAM50874 Cadherin EGF LAG s
33	20	0.7	20	24 ABB83213 G protein-coupled
34	20	0.7	20	24 ABB83214 G protein-coupled
35	20	0.7	3014	22 ABB68533 Human novel cytol
36	20	0.7	3014	22 AAU02196 Seven-pass transme
37	20	0.7	3014	24 ABB81979 Human GPCR CELSR1
38	20	0.7	3028	22 AAB08586 Human NOV7 protein
39	19	0.7	19	24 ABB83212 G protein-coupled
40	19	0.7	19	24 ABB83215 G protein-coupled
41	18	0.6	18	23 AAM50873 Cadherin EGF LAG s
42	18	0.6	18	24 ABB83216 G protein-coupled
43	18	0.6	874	18 AAM07600 Human G-protein re
44	18	0.6	874	20 AAM94075 Human G-protein co
45	18	0.6	884	19 AAM59666 Amino acid sequenc

## ALIGNMENTS

RESULT 1	AAU07054 standard; Protein; 2923 AA.
AC	AAU07054;
XX	
XX	24-OCT-2001 (first entry)
DT	
XX	Human Flamingo protein encoded by cDNA splice variant.
XX	
XX	Flamingo; human; splice variant; G-protein coupled receptor; diabetes;
KW	signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;
KW	anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;
KW	obesity; hypertension; hyperextension; urinary retention; angina pectoris;
KW	myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;
KW	benign prostatic hypertrophy; psychotic disorder; neurological disorder;
KW	manic depression; delirium; dementia; severe mental retardation;
KW	Huntington's disease; Gilles de la Tourette's syndrome; anorectic; anti-anorectic;
KW	antifungal; antiviral; antiprotozoal; anti-HIV; anorectic; anti-anorectic;
KW	antiparkinsonian; cardiac; anticholinergic; neuroprotective;
KW	antidepressant; anticonvulsant; antitense therapy; gene therapy.
OS	
XX	Homo sapiens.
XX	
PN	MO200161003-A1.
XX	
PD	23-AUG-2001.
XX	
PF	19-FEB-2001; 2001WO-GB00680.
XX	
PR	19-FEB-2000; 2000GB-0004196.
XX	

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX Testa TT;  
 XX WPI: 2001-502792/55.  
 DR N-PSDB; AAS11678.  
 XX  
 XX An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 XX  
 PS Claim 2; Page 33-34; 66pp; English.  
 XX  
 CC The sequence represents a human Flamingo polypeptide encoded by a cDNA  
 CC splice variant polynucleotide. Flamingo is a member of the G-protein  
 CC coupled receptor family, which is involved in signal transduction  
 CC pathways. By screening to identify compounds that stimulate or inhibit  
 CC the function or level of the protein, treatments can be developed for  
 CC various diseases and bacterial, fungal, protozoan and viral infections,  
 CC including HIV, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, angina pectoris, myocardial infarction, stroke,  
 CC ulcers, allergies and benign prostatic hypertrophy. Also treatable are  
 CC psychotic and neurological disorders such as anxiety, schizophrenia,  
 CC manic depression, delirium, dementia, severe mental retardation,  
 CC Huntington's disease and Gilles de la Tourette's syndrome.  
 CC  
 SQ Sequence 2923 AA;  
 Query Match 100.0%; Score 2923; DB 22; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVYVSAVDRDAHSVITVQITSGNTRNRSFITS 660  
 DB 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVYVSAVDRDAHSVITVQITSGNTRNRSFITS 660  
 QY 661 QSGGGVSLALPLDYLGLERQYLAATASTGTQDTQIYVNTVDANTHRPVQSSHYTN 720  
 DB 661 QSGGGVSLALPLDYLGLERQYLAATASTGTQDTQIYVNTVDANTHRPVQSSHYTN 720  
 QY 721 VNEDRPAGTTVVLISATDGTGNARITFEMEDSIPOFRIADDTGAVTVOAELDYEDQVS 780  
 DB 721 VNEDRPAGTTVVLISATDGTGNARITFEMEDSIPOFRIADDTGAVTVOAELDYEDQVS 780  
 QY 781 YTLAATARDNGIPQKSDTYTLBIYVNDVNDNAPOFLRDSYQGSVDEVPFTSVLQISAT 840  
 DB 781 YTLAATARDNGIPQKSDTYTLBIYVNDVNDNAPOFLRDSYQGSVDEVPFTSVLQISAT 840  
 QY 841 DRDSGLNGVFTVTPQGGDGDGDFVSTSGIVRTLRRLDRNVAYVLRAVADKGMPP 900  
 DB 841 DRDSGLNGVFTVTPQGGDGDGDFVSTSGIVRTLRRLDRNVAYVLRAVADKGMPP 900  
 QY 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVATDPDEGTNAQIMYQI 960  
 DB 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVATDPDEGTNAQIMYQI 960  
 QY 961 VEGNIPEVFQLDIPSGELTALVDLYEDRPEYVLVIQATSAPLVSATVHRLDRNDP 1020  
 DB 961 VEGNIPEVFQLDIPSGELTALVDLYEDRPEYVLVIQATSAPLVSATVHRLDRNDP 1020  
 QY 1021 PVLGNFELLFNNYVNNRSPFGGAI GRVADPDIDSLTYSFERNGLSLVTLNASTG 1080  
 DB 1021 PVLGNFELLFNNYVNNRSPFGGAI GRVADPDIDSLTYSFERNGLSLVTLNASTG 1080  
 QY 1081 ELKLSRALDNNRPLEAIMSVLSDGVHSTAQALRVITITTEMLHSTLRLDEMSPER 1140  
 DB 1081 ELKLSRALDNNRPLEAIMSVLSDGVHSTAQALRVITITTEMLHSTLRLDEMSPER 1140  
 QY 1141 FLSPILGLFTQAAVATLTPDPHVNVVFNQDTPADAGHILNVSLVSGPPGPGGPPPL 1200  
 DB 1141 FLSPILGLFTQAAVATLTPDPHVNVVFNQDTPADAGHILNVSLVSGPPGPGGPPPL 1200  
 QY 1201 PSEDLOERLYNRSLLTASAVRVLPEPDNICLRPCENYMCVSLRFDSSAPFASS 1260  
 DB 1201 PSEDLOERLYNRSLLTASAVRVLPEPDNICLRPCENYMCVSLRFDSSAPFASS 1260  
 QY 1261 VLFRRPIHPVGLRCRCPFGFTGDCYCEVDLYCYSRPCGPHGCRSREGGYTLCRDGYTG 1320  
 DB 1261 VLFRRPIHPVGLRCRCPFGFTGDCYCEVDLYCYSRPCGPHGCRSREGGYTLCRDGYTG 1320  
 QY 1321 EHCEVSARGRCCTPGVCNKGTCVNLVGGFKDCPSGDPPEKPYCQVTRSPFASFITF 1380  
 DB 1321 EHCEVSARGRCCTPGVCNKGTCVNLVGGFKDCPSGDPPEKPYCQVTRSPFASFITF 1380  
 QY 1381 RGLRORFHTLALSPFTEKRDGLLYNGRPNKHDVVALEVOVOLFPASGESITTVS 1440  
 DB 1381 RGLRORFHTLALSPFTEKRDGLLYNGRPNKHDVVALEVOVOLFPASGESITTVS 1440  
 QY 1441 PFVPGVSDGOMHTVQLKTYNKLQGTGLPOGSPBQKAAVTVVDCDGTVALRFGSVLG 1500  
 DB 1441 PFVPGVSDGOMHTVQLKTYNKLQGTGLPOGSPBQKAAVTVVDCDGTVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQSGSKSLDTGLPLLGVDLDESPFVRMRQFQCMKRNQJVSRIHDMAD 1560  
 DB 1501 NYSCAAQGTQSGSKSLDTGLPLLGVDLDESPFVRMRQFQCMKRNQJVSRIHDMAD 1560  
 QY 1561 FIANNGTVPGPCAKKVCDSNTCHNGTGVNOMDASCPCPLGFGKSCAQEAMNPOHL 1620  
 DB 1561 FIANNGTVPGPCAKKVCDSNTCHNGTGVNOMDASCPCPLGFGKSCAQEAMNPOHL 1620  
 QY 1621 GSSLVAMHGLSLPISQPTVLSLFRTRQADGVLLQAITRGSTITTLQREGHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISQPTVLSLFRTRQADGVLLQAITRGSTITTLQREGHVMSVEGT 1680  
 QY 1681 GLQASSLRLEPGRANDGMDHHAQALGASGGFGHAILSTFDYGGQABGNLGPRLHGLHS 1740



```

Db      1681  GLASSRLRLEPGANODMWHADLALGASGPGHALLSPYGOORABGNLPGRLHGLHLS 1740
Qy      1741  NITVGGI PGAGVAGFRGCLQGVRSVTPEGVNSIDPSHGESINVEQCSLPDPDSN 1800
Db      1741  NITVGGI PGAGVAGFRGCLQGVRSVTPEGVNSIDPSHGESINVEQCSLPDPDSN 1800
Qy      1801  PCGANVCSNDMOSYSSCDPGYVGNCTVNCOLNPEHOSVCTRKRSAHGTTCCEPPN 1860
Db      1801  PCGANVCSNDMOSYSSCDPGYVGNCTVNCOLNPEHOSVCTRKRSAHGTTCCEPPN 1860
Qy      1861  YLGPYCTRIDQCPFRGMWHPITCGPCNCDVSKGFPDCKNTSGEGCHCKENHTRPPGSP 1920
Db      1861  YLGPYCTRIDQCPFRGMWHPITCGPCNCDVSKGFPDCKNTSGEGCHCKENHTRPPGSP 1920
Qy      1921  CLLCDYPTGSLRVCDBEDQCPCKPGVI GRCDCRDNPAEVTNNGCEVNYDSCERAI 1980
Db      1921  CLLCDYPTGSLRVCDBEDQCPCKPGVI GRCDCRDNPAEVTNNGCEVNYDSCERAI 1980
Qy      1981  EAGIWMPTRRFGI PAAPCPKGSFGTAVRHCDHRGMLPNLNFCTISITSEIKGFAERL 2040
Db      1981  EAGIWMPTRRFGI PAAPCPKGSFGTAVRHCDHRGMLPNLNFCTISITSEIKGFAERL 2040
Qy      2041  QRNESGLDSGRSQQALLLNATQHTAGYFGSDVKVAYQLATRLAHSTORGEGLSATQ 2100
Db      2041  QRNESGLDSGRSQQALLLNATQHTAGYFGSDVKVAYQLATRLAHSTORGEGLSATQ 2100
Qy      2101  DVHFTENLRVGSALLDTANKRMHLLIQTGEGTAMLLQHYEAYASALANMBHTYLSPP 2160
Db      2101  DVHFTENLRVGSALLDTANKRMHLLIQTGEGTAMLLQHYEAYASALANMBHTYLSPP 2160
Qy      2161  TITPNTIVSVVLDKGNFAGALPRYEALRGEPPLLETTVLLPESVFETPPVVRPAG 2220
Db      2161  TITPNTIVSVVLDKGNFAGALPRYEALRGEPPLLETTVLLPESVFETPPVVRPAG 2220
Qy      2221  PGGAORPEELARQRHPELSQGEAVSVIITRTLAGLPHNDPDRGSLRVKRPINT 2280
Db      2221  PGGAORPEELARQRHPELSQGEAVSVIITRTLAGLPHNDPDRGSLRVKRPINT 2280
Qy      2281  PVSIVSVHDEBLLPRALDKPVTVOFRLTEBETKPI CVFMNHSILVSGVMSARGCE 2340
Db      2281  PVSIVSVHDEBLLPRALDKPVTVOFRLTEBETKPI CVFMNHSILVSGVMSARGCE 2340
Qy      2341  VVRNRSNHSVCCNMHTSPAVLMDVSRRENGELLPLKTLTYVALGTLTLPFPLTL 2400
Db      2341  VVRNRSNHSVCCNMHTSPAVLMDVSRRENGELLPLKTLTYVALGTLTLPFPLTL 2400
Qy      2401  LRILRSNQHGRIRNLTALGLAQVFLLAGINQADLPACTVIALLHFLYLCFESNALLE 2460
Db      2401  LRILRSNQHGRIRNLTALGLAQVFLLAGINQADLPACTVIALLHFLYLCFESNALLE 2460
Qy      2461  ALHLRYALTEVRDVTNGPMFYYMLGNGVPAFTTGLAVGLDPBGYGNPDFCWLSTYDTLI 2520
Db      2461  ALHLRYALTEVRDVTNGPMFYYMLGNGVPAFTTGLAVGLDPBGYGNPDFCWLSTYDTLI 2520
Qy      2521  WSRAGVAVRVSNSVFLYTLAARASCAHQRFKKGPSVGLQPSRAVLLLSATWLLAL 2580
Db      2521  WSRAGVAVRVSNSVFLYTLAARASCAHQRFKKGPSVGLQPSRAVLLLSATWLLAL 2580
Qy      2581  LSVNSDTLLFHYLPATCNCIQGFPIFLASYVVLSEKVRKALKACSRKSPDPALTTKSTL 2640
Db      2581  LSVNSDTLLFHYLPATCNCIQGFPIFLASYVVLSEKVRKALKACSRKSPDPALTTKSTL 2640
Qy      2641  TSSYNCPSPYADRLYQPYGDSAGSLHSTRSKSGQSYI PFLIRBSALNPGQPGGLG 2700
Db      2641  TSSYNCPSPYADRLYQPYGDSAGSLHSTRSKSGQSYI PFLIRBSALNPGQPGGLG 2700
Qy      2701  DPGSLLEGOQDHDPTDSLSLEDDOSGYASTHSDSEEBEERERKAAFGEGG 2760
Db      2701  DPGSLLEGOQDHDPTDSLSLEDDOSGYASTHSDSEEBEERERKAAFGEGG 2760
Qy      2761  WBSILPGARLPLHSTPKDGGPGKAPWPGDFTTAKSSNGAPBEERLRNENALSR 2820

```

```

Db      2761  WBSILPGARLPLHSTPKDGGPGKAPWPGDFTTAKSSNGAPBEERLRNENALSR 2820
Qy      2821  EBSLGLPSSAQPXKGLIKKCLPTISEKSLRLPLEQCTGSSRSASBEGSRGAPP 2880
Db      2821  EBSLGLPSSAQPXKGLIKKCLPTISEKSLRLPLEQCTGSSRSASBEGSRGAPP 2880
Qy      2881  RPPRQSLOELQNGWPIAMSITKAGTVNDDSSGSEPLFPNPLH 2923
Db      2881  RPPRQSLOELQNGWPIAMSITKAGTVNDDSSGSEPLFPNPLH 2923

RESULT 2
AAMS0866
ID AAMS0866 standard; Protein; 2923 AA.
XX
AC AAMS0866;
XX
DT 07-MAY-2002 (first entry)
XX
DE Cadherin EGF LAG seven-pass G-type receptor 2, basal cell marker.
XX
KW Cadherin EGF LAG seven-pass G-type receptor 2; receptor;
KW EGF-like domain, multiple 2; basal cell; marker; breast cancer;
KW tumour; diagnosis; prognosis; classification; human; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200208765-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US23843.
XX
PR 26-JUL-2000; 2000US-220967P.
XX
PA (STRD ) UNIV STANFORD.
PA (GENO-) APPLIED GENOMICS INC.
XX
PI Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
PI Seitz R;
XX
DR WPI; 2002-206094/26.
XX
PT Classifying and treating breast tumours associated with the expression
PT and activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14
PT and/or cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like
PT Domain, Multiple 2 -
XX
PS Claim 3; Fig 1C; 741pp; English.
XX
CC The present sequence is that of human cadherin EGF LAG seven-pass
CC G-type receptor 2, also known as EGF-like domain, multiple 2.
CC The invention provides new reagents and methods for the management
CC (e.g. detection, classification, provision of diagnostic and
CC prognostic information, treatment, etc.) of breast cancer. CDNA
CC microarray technology was used to identify genes whose expression
CC profile across a large group of tumour samples correlated with that
CC of cytokeratin 5 and cytokeratin 17, markers for basal cells of the
CC normal mammary lactation gland. Tumours that express cytokeratin
CC 5/6 and/or 17 have a poor prognosis relative to tumours overall.
CC The basal marker genes identified included those encoding cadherin
CC 3 (see AAMS0866), matrix metalloproteinase 14 (see AAMS0865) and
CC cadherin EGF LAG seven-pass G-type receptor 2. Detection of these
CC markers allows classification of a tumour sample as belonging to a
CC basal tumour subclass. By providing reagents that may reliably be
CC used to classify tumours as belonging to a basal subclass, the
CC invention enables a variety of methods for improving therapeutic
CC options for patients with breast cancer and for individualising
CC therapy. Potential therapeutic agents include compounds that
CC modulate breast basal cell marker genes or that modulate
CC polypeptides encoded by these genes. In particular, therapeutic
CC agents include antibodies that specifically bind to the basal
CC marker proteins and which recognise basal cells of normal mammary

```



Db	1981	PAGIWMPTRTGRLPAAAPCPKSGPETAARHCHDHNGMLPMLFNCTSTIPBEIKGFAPRL	204
Qy	2041	QRNPSGLDSGSSQQLALLRNATQHTAGYFGSDVAVAYQALTRLIAHSTORGFGLSATQ	2100
Db	2041	QRNPSGLDSGSSQQLALLRNATQHTAGYFGSDVAVAYQALTRLIAHSTORGFGLSATQ	2100
Qy	2101	DVHFNENLIRVGSALLDTANKRMHELIOOTRGCTWMLQHYEAVASALAQNWRHTYLSPF	216C
Db	2101	DVHFNENLIRVGSALLDTANKRMHELIOOTRGCTWMLQHYEAVASALAQNWRHTYLSPF	216G
Qy	2161	TIIVPNIVISVVRLDKGNFAGAKLPRYALRGSGPPDETTYIIPESYFRETPIVVRPAG	222C
Db	2161	TIIVPNIVISVVRLDKGNFAGAKLPRYALRGSGPPDETTYIIPESYFRETPIVVRPAG	222G
Qy	2221	PGEAOPEELARORRHPELSQGEAVASVIIYRTLAGLLPHNYDPDKSLRVPKRP1INT	228G
Db	2221	PGEAOPEELARORRHPELSQGEAVASVIIYRTLAGLLPHNYDPDKSLRVPKRP1INT	228G
Qy	2281	PVSVISVYHDBEELLPRALDKPVYVQFLLIFTEERKPICVWNHNSILVSGGMSARCE	234G
Db	2281	PVSVISVYHDBEELLPRALDKPVYVQFLLIFTEERKPICVWNHNSILVSGGMSARCE	234G
Qy	2341	VVFNRESHVSQCCNMHTSFALVMDVSRRENEIIPKLITVYALGVTALALLTFEFLT	240G
Db	2341	VVFNRESHVSQCCNMHTSFALVMDVSRRENEIIPKLITVYALGVTALALLTFEFLT	240G
Qy	2401	LRIILRSNHGIRNULTAALGALVFLGIGNQADLPFACVTYALILHFLYLCFSWALLE	246G
Db	2401	LRIILRSNHGIRNULTAALGALVFLGIGNQADLPFACVTYALILHFLYLCFSWALLE	246G
Qy	2461	ALHLYRALTEVRDVTNGTGMRYTMLGKCVAPFITGLAVGLDPEGYGNDFCWLSTYDTLI	252G
Db	2461	ALHLYRALTEVRDVTNGTGMRYTMLGKCVAPFITGLAVGLDPEGYGNDFCWLSTYDTLI	252G
Qy	2521	WSPAGPVAFVMSVFLYTLIARASCAARQGFKEKGVSGLOPSFALLILSTWTLAL	258G
Db	2521	WSPAGPVAFVMSVFLYTLIARASCAARQGFKEKGVSGLOPSFALLILSTWTLAL	258G
Qy	2581	LSVNSDILLFHYLPATCNCIOGPFIFLSYVLSKEVRKALMLACSRKSPDPALTTKSTL	264G
Db	2581	LSVNSDILLFHYLPATCNCIOGPFIFLSYVLSKEVRKALMLACSRKSPDPALTTKSTL	264G
Qy	2641	TSSINCPSPYADGRUYOPYGDSAGSLHSTSRGKSQPSYIPFLIRRESALNPQGGPGLG	270G
Db	2641	TSSINCPSPYADGRUYOPYGDSAGSLHSTSRGKSQPSYIPFLIRRESALNPQGGPGLG	270G
Qy	2701	DPGSLPIEGCQOQOHDPTDSDSLSLHDDQSGSYASTSSDPSSEEESEEESEEAAPRERG	276G
Db	2701	DPGSLPIEGCQOQOHDPTDSDSDSLSLHDDQSGSYASTSSDPSSEEESEEESEEAAPRERG	276G
Qy	2761	WDSLILGPGAEERLPHSTPTKDGPGPGKAPMGDFGTTAKSSGNGCAPERLRRENGDALSR	282G
Db	2761	WDSLILGPGAEERLPHSTPTKDGPGPGKAPMGDFGTTAKSSGNGCAPERLRRENGDALSR	282G
Qy	2821	EGSISGLPIPGSSAOPHKGLIKKKCLPTISEKSLRLPLEQCTGSSRGSSASBGRGGBP	288G
Db	2821	EGSISGLPIPGSSAOPHKGLIKKKCLPTISEKSLRLPLEQCTGSSRGSSASBGRGGBP	288G
Qy	2881	RPPRROSLQEBOLNGWMPIMASIKAGTUDEDSGSEFLPFNLH 2923	
Db	2881	RPPRROSLQEBOLNGWMPIMASIKAGTUDEDSGSEFLPFNLH 2923	
RESULT 3			
ID	ABP82018	standard; Protein: 2923 AA.	
XX	ABP82018;		
XX	AC		
XX	04-MAR-2003	(first entry)	
XX	Human GPCR CELSR2 protein SEQ ID NO:524.		
XX			

KV	G protein-coupled receptor; GPCR antigenic peptide; gene therapy;
KW	G protein-coupled receptor modulator; antibody; immune-related disease;
KX	growth-related disease; cell regeneration-related disease; AIDS; cancer;
KY	immunological-related cell proliferative disease; autoimmune disease;
LK	Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
LM	osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KM	grief versus host disease; Parkinson's disease; multiple sclerosis; pain;
KN	psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KP	mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KQ	hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KR	ulcer.
KS	Homo sapiens.
KT	WO20061087-A2.
KU	08-AUG-2002.
KV	19-DEC-2001; 2001WO-US50107.
KW	19-DEC-2000; 2000US-257144P.
KX	(LIFE-) LIFESPAN BIOSCIENCES INC.
KY	Burner GC, Roush CL, Brown JP;
LA	WPI; 2003-046718/04.
LB	N-PSDB; ABZ42868.
LC	New isolated antigenic peptides e.g., for G protein-coupled receptors
LD	(GPCR), useful for diagnosing and designing drugs for treating
LE	conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
LF	cancer or autoimmune diseases -
LG	Disclosure; Fig 1; 523pp; English.
LH	The present invention describes antigenic peptides (1) comprising:
LI	(a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
LJ	G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
LK	and (2) an isolated antibody having high specificity and high affinity
LL	or avidity for a particular GPCR. (1) can be used as GPCR modulators and
LM	in gene therapy. The antigenic peptides for GPCRs are useful in detecting
LN	an antibody against a particular GPCR, and in the production of specific
LO	antibodies. The peptides and antibodies are also useful for detecting the
LP	presence or absence of corresponding GPCRs. The antigenic peptides for
LQ	GPCRs and antibodies are useful for diagnosing and designing drugs for
LR	treating immune-related diseases, growth-related diseases, cell
LS	regeneration-related disease, immunological-related cell proliferative
LT	diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
LU	atherosclerosis, bacterial, fungal, protozoan or viral infections,
LV	osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
LW	inflammation, allergies, Crohn's disease, diabetes, graft versus host
LX	disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
LY	anxiety, depression, schizophrenia, dementia, mental retardation, memory
MK	loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, memory
ML	hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
MM	any other disorder in which GPCRs are involved. The antibodies may be
MN	used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
MO	GPCR proteins given in ABP81675 to ABP82018, which are used in the
MP	exemplification of the present invention.
MQ	Sequence 2923 AA;
MR	Query Match 100.0%; Score 2923; DB 24; Length 2923;
MS	Best Local Similarity 100.0%; Pred. No. 0;
MT	Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
MY	1 MRSPATVPTPPPLLLLLLLLPPPLGDDVGPCKSLGRSGSSGACAPMGWLCP 60
DQ	1 MRSATVPTPPPLLLLLLLLPPPLGDDVGPCKSLGRSGSSGACAPMGWLCP 60
EY	61 SASNIWLTYSRCADAGTGLTGHLVPHHDGLRWVCPSSEBAHILPLPAPEGWCSSLIG 120

Dh	61	SASINMLTSTRCRAGTELTHGLVPHHDLRWCPBESBAHPLPAPAPGCGWSCRLLGIG	120
Qy	121	GHLSPOGKLTLPBEPHCLKAPRLRCQCKLAQAPGLRAGERSPEPSLGRKRNNTAPQ	180
Dh	121	GHLSPOGKLTLPBEPHCLKAPRLRCQCKLAQAPGLRAGERSPEPSLGRKRNNTAPQ	180
Qy	181	POPSYQATVBPENOPAGTPVASLRADIDEGBAGLEXTMDALPDSRNOFSLDPVTGA	240
Dh	181	POPSYQATVBPENOPAGTPVASLRADIDEGBAGLEXTMDALPDSRNOFSLDPVTGA	240
Qy	241	VTTBEBLRETKSTHVFRTVTAQDHMPRSALATLTITVTDNDHPVPEQOEKESLRE	300
Dh	241	VTTBEBLRETKSTHVFRTVTAQDHMPRSALATLTITVTDNDHPVPEQOEKESLRE	300
Qy	301	NLEVEGYEVLTVRATVDGAPPNANILYRLLEGSGSPSEVEFIDPRSGVIRTRGVNDEEV	360
Dh	301	NLEVEGYEVLTVRATVDGAPPNANILYRLLEGSGSPSEVEFIDPRSGVIRTRGVNDEEV	360
Qy	361	ESYQUTVEASDQGDGPGPRSTTAAVFLSVEDNDNAPQFSBKRYVQVREDVTPGAPYLR	420
Dh	361	ESYQUTVEASDQGDGPGPRSTTAAVFLSVEDNDNAPQFSBKRYVQVREDVTPGAPYLR	420
Qy	421	VTASDRKGSNAVVHYSIMSGNARGQFYLDAGALDVSPLDYETKEYTLRVADQDG	480
Dh	421	VTASDRKGSNAVVHYSIMSGNARGQFYLDAGALDVSPLDYETKEYTLRVADQDG	480
Qy	481	RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADAGDNARL	540
Dh	481	RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADAGDNARL	540
Qy	541	EYRLAGVGHDPFPTINNGTGMISVAABLDREEDVDFSGVEARHDGHPALASASVTV	600
Dh	541	EYRLAGVGHDPFPTINNGTGMISVAABLDREEDVDFSGVEARHDGHPALASASVTV	600
Qy	601	LDVANDNPTFQPEYTVRLNEDAAVGTSVTVSAVVDDBASVITVQITSGTRRPSITS	660
Dh	601	LDVANDNPTFQPEYTVRLNEDAAVGTSVTVSAVVDDBASVITVQITSGTRRPSITS	660
Qy	661	QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVNTDANTHRPVFGSHYTVN	720
Dh	661	QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVNTDANTHRPVFGSHYTVN	720
Qy	721	VNERPAGTGYVLLISADDEBTGENARITTFMEDSIPQFRIDADTGAVTTQAELEDEYQVS	780
Dh	721	VNERPAGTGYVLLISADDEBTGENARITTFMEDSIPQFRIDADTGAVTTQAELEDEYQVS	780
Qy	781	YTLAITARDNGIPKQSDPTTYLEILVNDVNDNAPOFLRDSYQGSYVEDVPPTSVLQISAT	840
Dh	781	YTLAITARDNGIPKQSDPTTYLEILVNDVNDNAPOFLRDSYQGSYVEDVPPTSVLQISAT	840
Qy	841	DRDSGLNGRVEYTFQGGDGDGDFIVESTSGIVRTLRLDRENVAAQYVLAAYDKMPP	900
Dh	841	DRDSGLNGRVEYTFQGGDGDGDFIVESTSGIVRTLRLDRENVAAQYVLAAYDKMPP	900
Qy	901	ARTMEYTVTVLVNDVNDPVPEKDEPFVPEENSPIGLAVARVATTPDBSTNAQIMYQI	960
Dh	901	ARTMEYTVTVLVNDVNDPVPEKDEPFVPEENSPIGLAVARVATTPDBSTNAQIMYQI	960
Qy	961	VEGNIPEVFOLDIPSGELTALVDLYEDRPEYVVIQNTSAPLYSRATVHRLDRDNP	1020
Dh	961	VEGNIPEVFOLDIPSGELTALVDLYEDRPEYVVIQNTSAPLYSRATVHRLDRDNP	1020
Qy	1021	PVLGNFELFNNTVNNRSSSPFGAIGRPVADPDIDSILYSPERGNELSLVLANASTG	1080
Dh	1021	PVLGNFELFNNTVNNRSSSPFGAIGRPVADPDIDSILYSPERGNELSLVLANASTG	1080
Qy	1081	ELKSRALDNNRPLEAITSVYSDGHSVTAQCLARTIITDMLTSTILRLDMSPER	1140
Dh	1081	ELKSRALDNNRPLEAITSVYSDGHSVTAQCLARTIITDMLTSTILRLDMSPER	1140
Qy	1141	FLSPLGLFIOAVAAATLTPPDHVVFNVOGDTAPGSHILANVSLVGOPGPGGPPFL	1200
Dh	1141	FLSPLGLFIOAVAAATLTPPDHVVFNVOGDTAPGSHILANVSLVGOPGPGGPPFL	1200
Dh	1201	PSEDLOERLYLNSLTLTAISOAVLPEPDNI CLREPCENMRCSVLRPSSAPFLSSS	1260
Qy	1201	PSEDLOERLYLNSLTLTAISOAVLPEPDNI CLREPCENMRCSVLRPSSAPFLSSS	1260
Dh	1261	VLFRPIHPVGLRCRCBPGEFGDYCETEVYLCYSRPCGPHRCRSRGGYTCLCRDGYTG	1320
Qy	1261	VLFRPIHPVGLRCRCBPGEFGDYCETEVYLCYSRPCGPHRCRSRGGYTCLCRDGYTG	1320
Dh	1321	EHCEVSARSRCRTPGVCKNGTCVNLVGGFKDCBPGDEKPYCQVTTSSFPASHPTF	1380
Qy	1321	EHCEVSARSRCRTPGVCKNGTCVNLVGGFKDCBPGDEKPYCQVTTSSFPASHPTF	1380
Dh	1381	RGLRORPHFTALSPATKERDGLLYNGRFNEKHDFAVLEVIOBOVOLTFSAGSTTVS	1440
Qy	1381	RGLRORPHFTALSPATKERDGLLYNGRFNEKHDFAVLEVIOBOVOLTFSAGSTTVS	1440
Dh	1441	PFVPGVSDGQMTVOLKTYNKPDLQYGLPQSPSOKVAVTVTDGDTVALRFGSVLG	1500
Qy	1441	PFVPGVSDGQMTVOLKTYNKPDLQYGLPQSPSOKVAVTVTDGDTVALRFGSVLG	1500
Dh	1501	NYSCAOGTQGSKSLDLTGPIILGVPDLPSFPVRMQPVGCMRNLOVDSRHIDMAD	1560
Qy	1501	NYSCAOGTQGSKSLDLTGPIILGVPDLPSFPVRMQPVGCMRNLOVDSRHIDMAD	1560
Dh	1561	FIANNGTVPCCPAKKVNCNSNTCHNGGTCVNOQDASCECPFLFGGSCAQOEMANPOHFL	1620
Qy	1561	FIANNGTVPCCPAKKVNCNSNTCHNGGTCVNOQDASCECPFLFGGSCAQOEMANPOHFL	1620
Dh	1621	GSSLVAMHGLSLPISQPMYLSIMFRTROADGVLLQAITRGRSTTTQLRREGVMLSVETG	1680
Qy	1621	GSSLVAMHGLSLPISQPMYLSIMFRTROADGVLLQAITRGRSTTTQLRREGVMLSVETG	1680
Dh	1681	GLQASSRLBEPGANQDMHQAOLGASGSPGHALISFPYGGQRBAGNLPRLHGLHS	1740
Qy	1681	GLQASSRLBEPGANQDMHQAOLGASGSPGHALISFPYGGQRBAGNLPRLHGLHS	1740
Dh	1741	NITVGGIPGAGVARGFRGCLQGVAVSDTPBGSNLDPSHGHSINVBQCSLPDQDSN	1800
Qy	1741	NITVGGIPGAGVARGFRGCLQGVAVSDTPBGSNLDPSHGHSINVBQCSLPDQDSN	1800
Dh	1801	PCPANSYCSNDMSYSCSDPGYGNCTNVCNLANFCHOSVTRKPSAPHGTTCECPFN	1860
Qy	1801	PCPANSYCSNDMSYSCSDPGYGNCTNVCNLANFCHOSVTRKPSAPHGTTCECPFN	1860
Dh	1861	YLGPFYCESTRIDQCPRGMMGHPTCGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSPF	1920
Qy	1861	YLGPFYCESTRIDQCPRGMMGHPTCGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSPF	1920
Dh	1921	CLLDCYPTGSLRVCDPBEDQCPCKRGVYIGRQCDRCNPPFAEVTNNGEVNYSQPRAI	1980
Qy	1921	CLLDCYPTGSLRVCDPBEDQCPCKRGVYIGRQCDRCNPPFAEVTNNGEVNYSQPRAI	1980
Dh	1981	EAGIMWPRTRFGILPAAAPCPKSGFTAVRHCDHRMULPNNLPNCSTITPSELKGAEBL	2040
Qy	1981	EAGIMWPRTRFGILPAAAPCPKSGFTAVRHCDHRMULPNNLPNCSTITPSELKGAEBL	2040
Dh	2041	QRNBSGLDGSRSQALLRNATORTAGYFGSPDKVAYOLAATRLAHESITORFGFSATQ	2100
Qy	2041	QRNBSGLDGSRSQALLRNATORTAGYFGSPDKVAYOLAATRLAHESITORFGFSATQ	2100
Dh	2101	DVAFTEMLRVSGALLDTANKRMELIQTEGGTAWLQHYEAYASALAQNMRHTYLSPF	2160
Qy	2101	DVAFTEMLRVSGALLDTANKRMELIQTEGGTAWLQHYEAYASALAQNMRHTYLSPF	2160
Dh	2161	TIYTPNIVISVVRLDGNFAGAKLPYRYEALRGOPDLTETVILPESVRETPPVVRPAG	2220
Qy	2161	TIYTPNIVISVVRLDGNFAGAKLPYRYEALRGOPDLTETVILPESVRETPPVVRPAG	2220
Dh	2221	PGEAQEPBELLARQRHPELSGGEAVASVITRTTAGLLPHNYDPDKSLRVPKRPDIINT	2280
Qy	2221	PGEAQEPBELLARQRHPELSGGEAVASVITRTTAGLLPHNYDPDKSLRVPKRPDIINT	2280
Dh	2281	PGEAQEPBELLARQRHPELSGGEAVASVITRTTAGLLPHNYDPDKSLRVPKRPDIINT	2280

```

QY 2281 PVVSIIVHDEHLLPRLADKPVTVOFRLLETERTPICVPMNHSLIVSGTGMARCE 2340
DB 2281 PVVSIIVHDEHLLPRLADKPVTVOFRLLETERTPICVPMNHSLIVSGTGMARCE 2340
QY 2341 VVERNEHSHVSCCNHMTSPFAVMDVSRRENGEILPKTLTYVALAGTLAALLTFPELTL 2400
DB 2341 VVERNEHSHVSCCNHMTSPFAVMDVSRRENGEILPKTLTYVALAGTLAALLTFPELTL 2400
QY 2401 LKILSNQGIIRNLTALAGLAQVFLGLINQADLPACTVIAILHFLYLCFPMALL 2460
DB 2401 LKILSNQGIIRNLTALAGLAQVFLGLINQADLPACTVIAILHFLYLCFPMALL 2460
QY 2461 AALVRLALREVDVNTGMPFYMGMGPARTTGLAVGLDDEGYNPCMLSTDTLI 2520
DB 2461 AALVRLALREVDVNTGMPFYMGMGPARTTGLAVGLDDEGYNPCMLSTDTLI 2520
QY 2521 WSPAGVAVAVMSVFLYLIAAASCAARQCEKGPVSGLOPFAVILLISATWLLAL 2580
DB 2521 WSPAGVAVAVMSVFLYLIAAASCAARQCEKGPVSGLOPFAVILLISATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFATCNCTIQGPFIPLSYVLISKEVRKALKACSRKSPDPALTTKSTL 2640
DB 2581 LSVNSDTLLFHYLFATCNCTIQGPFIPLSYVLISKEVRKALKACSRKSPDPALTTKSTL 2640
QY 2641 TSSNCPSPYADGRLYQPYGDSAGLSHSTRSGKSPSYIPLLRRESALNPGQGPGLG 2700
DB 2641 TSSNCPSPYADGRLYQPYGDSAGLSHSTRSGKSPSYIPLLRRESALNPGQGPGLG 2700
QY 2701 DPGSLFLBGOQDQHPDPTSDSLSLIEDQSGSYASTHSDSEBEEBEEBAAFPBEG 2760
DB 2701 DPGSLFLBGOQDQHPDPTSDSLSLIEDQSGSYASTHSDSEBEEBEEBAAFPBEG 2760
QY 2761 WDSLLGPGAEPLHSTPPOGPGPGKADPGDFTTAAESSGNGAPBEERLENDALSR 2820
DB 2761 WDSLLGPGAEPLHSTPPOGPGPGKADPGDFTTAAESSGNGAPBEERLENDALSR 2820
QY 2821 BBSLPLPSSSAQPHKGLIKKCLPTISKSSILRLPLEQCTGSSRGSSASBGSRGPPP 2880
DB 2821 BBSLPLPSSSAQPHKGLIKKCLPTISKSSILRLPLEQCTGSSRGSSASBGSRGPPP 2880
QY 2881 RPPRQSLQEQNLGVMPIAMSIKAGTVDESSGSEFLFPNPLH 2923
DB 2881 RPPRQSLQEQNLGVMPIAMSIKAGTVDESSGSEFLFPNPLH 2923
QY 2923 RPPRQSLQEQNLGVMPIAMSIKAGTVDESSGSEFLFPNPLH 2923
DB 2923 RPPRQSLQEQNLGVMPIAMSIKAGTVDESSGSEFLFPNPLH 2923

```

```

PR 28-MAR-2001; 2001US-279619P.
PR 29-MAR-2001; 2001US-280067P.
PR 29-MAR-2001; 2001US-280068P.
PR 16-MAY-2001; 2001US-291280P.
PR 17-MAY-2001; 2001US-291829P.
PR 17-MAY-2001; 2001US-291849P.
PR 19-JUN-2001; 2001US-299428P.
PR 20-JUN-2001; 2001US-299776P.
PR 20-JUN-2001; 2001US-300001P.
PA (INCYTE GENOMICS INC.
XX
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J,
PI Dutoir GE, Hillman JL, Yu JY, Tnaon O, Yap PE, Amshley SR,
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH,
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,
PI Flores V, Marwaha R, Lo A, Ian RV, Urashka ME;
XX
DR WPI: 2003-058431/05.
XX N-PSDB; ABX34546.
XX
PT New purified disease detection and treatment molecule proteins and
PT polynucleotides, useful for diagnosing, treating or preventing cancers
PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
PT or hepatitis
XX
PS Claim 27; SEQ ID NO 503; 339pp + Sequence Listing; English.
XX
CC This invention describes a novel disease detection and treatment molecule
CC polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,
CC osteoprotic, cytoprotic, anti-HIV, haemostatic, nephrotropic,
CC antianemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC and the polypeptides of the invention can be used for gene therapy,
CC protein replacement therapy and are useful for treating a variety of
CC diseases or conditions. These polypeptides or polynucleotides are
CC particularly useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
CC diseases, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC hepatitis. ABU1450-ABU1845 represent the MDPT polynucleotides encoded
CC by ABU1450-ABU1845, described in the disclosure of the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2894 AA;

```

```

Query Match 99.0%; Score 2894; DB 24; Length 2894;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 30 LGDVGPCRSIGSGRGSGGACAPMGWLCPSASNLMTYSCRDAGTGLVPHHDG 89
DB 1 LGDVGPCRSIGSGRGSGGACAPMGWLCPSASNLMTYSCRDAGTGLVPHHDG 60
QY 90 LRWCPSSEAHIPPPABEGCPMSCRLLIGIGHLSPQKTLPPSHPCIKAPRLRCQSC 149
DB 61 LRWCPSSEAHIPPPABEGCPMSCRLLIGIGHLSPQKTLPPSHPCIKAPRLRCQSC 120
QY 150 LQAQPLRAGRSRSEBSIGRRKRNVTAPQPPPSYATYBENOPACTPVASLRALPD 209
DB 121 LQAQPLRAGRSRSEBSIGRRKRNVTAPQPPPSYATYBENOPACTPVASLRALPD 180
QY 210 EGEAGRLTYMDALFDSRSNQPFLSDPYTGAVTAAEELDETKSGTHVRAVTAODHGMRR 269
DB 181 EGEAGRLTYMDALFDSRSNQPFLSDPYTGAVTAAEELDETKSGTHVRAVTAODHGMRR 240
QY 270 SALATLTLVTDNDHPVFEQOEYKESLRENLEVGFEVLTVRAITDGDAPPNNATLYLL 329
DB 241 SALATLTLVTDNDHPVFEQOEYKESLRENLEVGFEVLTVRAITDGDAPPNNATLYLL 300
QY 330 EGSQSPSEVFEIDPRSGVIRTRGPVDRBEVSYQLVTEASDQGRDPGRSTTAAVFLSV 389

```

Db 301 EGSGSGSEFEIEDPRSGVIRTRGPDVIREEVESYQLTVEASDQGRDPRSTTAVELSV 360  
Qy 390 EDDNDNAPORSEKKRYVQVREDVTPGA PYLAVTASDSDKSNAYVHSINSNGNRGQFYL 449  
Db 361 EDDNDNAPORSEKKRYVQVREDVTPGA PYLAVTASDSDKSNAYVHSINSNGNRGQFYL 420  
Qy 450 DAQGTALDVNSPLDYETTKETKYLVRADQDGRPELNSVSGLYTVQVLDINDNAPIFVSTP 509  
Db 421 DAQGTALDVNSPLDYETTKETKYLVRADQDGRPELNSVSGLYTVQVLDINDNAPIFVSTP 480  
Qy 510 PQATVLESVPVGTIVLHVQALDADAGDNARLEFYLAVGHDPPPTINNGTGMISVAEELD 569  
Db 481 PQATVLESVPVGTIVLHVQALDADAGDNARLEFYLAVGHDPPPTINNGTGMISVAEELD 540  
Qy 570 REEDVFSFGVEARDHGTALTAASVSVTVLDVNDNPPFTOPEYTVRLNEDAAVGTSV 629  
Db 541 REEDVFSFGVEARDHGTALTAASVSVTVLDVNDNPPFTOPEYTVRLNEDAAVGTSV 600  
Qy 630 VTVSAVDRDAHSVITTYQITSGNTNRPSITSGSGGLVSLALPLDYKLERQYVLAATASD 689  
Db 601 VTVSAVDRDAHSVITTYQITSGNTNRPSITSGSGGLVSLALPLDYKLERQYVLAATASD 660  
Qy 690 GTRDQTAQIVNVVTDANTHRPVFQSSHYTVANVEDRPACTTVTLISATDEPTGEMARITY 749  
Db 661 GTRDQTAQIVNVVTDANTHRPVFQSSHYTVANVEDRPACTTVTLISATDEPTGEMARITY 720  
Qy 750 FMEDSIPQFRIDATGAVTTOAELDYEDQVSTYLAITARDNGIPQKSDTTYLETIVNDVN 809  
Db 721 FMEDSIPQFRIDATGAVTTOAELDYEDQVSTYLAITARDNGIPQKSDTTYLETIVNDVN 780  
Qy 810 DNABQPLRDSYQGSVYEDVPPTSVLQISATDRDSGLNGRVYTFQGGDDGDGDFIVEST 869  
Db 781 DNABQPLRDSYQGSVYEDVPPTSVLQISATDRDSGLNGRVYTFQGGDDGDGDFIVEST 840  
Qy 870 SGIVRTLRRLDRENVAAQVTLBAAYAVDKMPPARTPMEYTVTVLDVNDNPPFREDDEVDV 929  
Db 841 SGIVRTLRRLDRENVAAQVTLBAAYAVDKMPPARTPMEYTVTVLDVNDNPPFREDDEVDV 900  
Qy 930 VEENSPIGLAVARVATDPDEGTNAQIMYQIVEGNIPEVQOLDIFSGELTALVOLDEDYR 989  
Db 901 VEENSPIGLAVARVATDPDEGTNAQIMYQIVEGNIPEVQOLDIFSGELTALVOLDEDYR 960  
Qy 990 PEYVAVIQAISAPVLSRAVTHVRLLDNRDNPVLAGNEFILFNNTYVNRSSSPGCAIGRV 1049  
Db 961 PEYVAVIQAISAPVLSRAVTHVRLLDNRDNPVLAGNEFILFNNTYVNRSSSPGCAIGRV 1020  
Qy 1050 PAHDPOISDSITVSPERGENELSTVILNASNGELKLSRALDNNRPLEAMSTLVSDGVHSV 1109  
Db 1021 PAHDPOISDSITVSPERGENELSTVILNASNGELKLSRALDNNRPLEAMSTLVSDGVHSV 1080  
Qy 1110 TAQCALRVTIITDEMLTHTSITLRLSDMSPERFLSPILGLFIQAVAAATLATPPDVVFNV 1169  
Db 1081 TAQCALRVTIITDEMLTHTSITLRLSDMSPERFLSPILGLFIQAVAAATLATPPDVVFNV 1140  
Qy 1170 QRDTPADPGHILNLSVSGPPGPGGPPPLPSEDLQERLYLNLSLTAISAQVRLPDD 1229  
Db 1141 QRDTPADPGHILNLSVSGPPGPGGPPPLPSEDLQERLYLNLSLTAISAQVRLPDD 1200  
Qy 1230 NICIREPENMRCVSVLRFPSSAPFIASSSVLRPHHPVGLRCRCRPGTGYCEREV 1289  
Db 1201 NICIREPENMRCVSVLRFPSSAPFIASSSVLRPHHPVGLRCRCRPGTGYCEREV 1260  
Qy 1290 DLCTSRPQPGHRCRSGEGYTCLCRDGYTGHCHCVSARSRCRTPGVCNKGATCVNLVVG 1349  
Db 1261 DLCTSRPQPGHRCRSGEGYTCLCRDGYTGHCHCVSARSRCRTPGVCNKGATCVNLVVG 1320  
Qy 1350 GFKCDCEBGBDEKPYCOVTTSSFPAAHSTITTRGRLFORHFTALSFATKERDGLLNGR 1409  
Db 1321 GFKCDCEBGBDEKPYCOVTTSSFPAAHSTITTRGRLFORHFTALSFATKERDGLLNGR 1380  
Qy 1410 FNEKHDPALEVIQEOVOLTSAGSSTTVSPVVGVSDDQOMHVOLKYNKPLAGTGT 1469

Db 1381 FNEKHDPALEVIQEOVOLTSAGSSTTVSPVVGVSDDQOMHVOLKYNKPLAGTGT 1440  
Qy 1470 LPQSPBOKVAVTVYDGCOTGVALRFGSVLGNVSCAAOGTQGGSKSLDTGPLLGGVP 1529  
Db 1441 LPQSPBOKVAVTVYDGCOTGVALRFGSVLGNVSCAAOGTQGGSKSLDTGPLLGGVP 1500  
Qy 1530 DLBESFPVNRQOFVGCNRNLOVDSRHI DMADFIANNGTVPGCBAKXNCDNSNCHNGTIC 1589  
Db 1501 DLBESFPVNRQOFVGCNRNLOVDSRHI DMADFIANNGTVPGCBAKXNCDNSNCHNGTIC 1560  
Qy 1590 VNOQWDAFSCBCEPLGREGKSCAQEMANPOHFLGSSLYAHNGLSLPIQOPWTLSTMFTRQA 1649  
Db 1561 VNOQWDAFSCBCEPLGREGKSCAQEMANPOHFLGSSLYAHNGLSLPIQOPWTLSTMFTRQA 1620  
Qy 1650 DGVILQAITRGRSTITLQIREGVMLSVGTGLOAASSLRLEPGRANDGMVHQAQLGAS 1709  
Db 1621 DGVILQAITRGRSTITLQIREGVMLSVGTGLOAASSLRLEPGRANDGMVHQAQLGAS 1680  
Qy 1710 GGFBAHILSPDYQOQABGNLGRHLGHLNITVGGIPQAGVARGPGLQGVAVSD 1769  
Db 1681 GGFBAHILSPDYQOQABGNLGRHLGHLNITVGGIPQAGVARGPGLQGVAVSD 1740  
Qy 1770 TPEGVNSLDPSSHGHSINVEQGCSSLPPDCSNPCPANSYCSNDMDYSGCCDPRGYDNDCT 1829  
Db 1741 TPEGVNSLDPSSHGHSINVEQGCSSLPPDCSNPCPANSYCSNDMDYSGCCDPRGYDNDCT 1800  
Qy 1830 NVCDLNPCEHQSCTKRPSPAHGYTCECPNLYGPFCEFRIDOPCRPGMMGHPTGCPNC 1889  
Db 1801 NVCDLNPCEHQSCTKRPSPAHGYTCECPNLYGPFCEFRIDOPCRPGMMGHPTGCPNC 1860  
Qy 1890 DVSQKDFDPCNKTSGBCHCKENHRRPGSPFTCLCDCTPGSISRYCDPBGQCPCKPGV 1949  
Db 1861 DVSQKDFDPCNKTSGBCHCKENHRRPGSPFTCLCDCTPGSISRYCDPBGQCPCKPGV 1920  
Qy 1950 IGRQCRCONPFAEYTVNGCEVNVYDSCPRALTEGIMPPTRFGLPAAAPCPKSPGTAVR 2009  
Db 1921 IGRQCRCONPFAEYTVNGCEVNVYDSCPRALTEGIMPPTRFGLPAAAPCPKSPGTAVR 1980  
Qy 2010 HCEBHRGMLPBNLFNCTSIITFSELKGAERLQBNESGLDGRSQOULALLRNATOHTAGY 2069  
Db 1981 HCEBHRGMLPBNLFNCTSIITFSELKGAERLQBNESGLDGRSQOULALLRNATOHTAGY 2040  
Qy 2070 FGSQDVAVAYOLATRLLAHSTQRFGLSATODVHFTBNLIRVGSALLDTANKHWEILQ 2129  
Db 2041 FGSQDVAVAYOLATRLLAHSTQRFGLSATODVHFTBNLIRVGSALLDTANKHWEILQ 2100  
Qy 2130 TBGSTMLOHRYEAYASALAONNRHTYLSPTITVTENIYISVVRLDKGNPAGAKLPRYEA 2189  
Db 2101 TBGSTMLOHRYEAYASALAONNRHTYLSPTITVTENIYISVVRLDKGNPAGAKLPRYEA 2160  
Qy 2190 LRGSOPPDLETTYILPESVFERETPPVVRPAGPBAQEPBELARQRRHPELSQGBAVASY 2249  
Db 2161 LRGSOPPDLETTYILPESVFERETPPVVRPAGPBAQEPBELARQRRHPELSQGBAVASY 2220  
Qy 2250 IYRTTLAQLPHNYDDPKSLRYPKXPIINTPVASISVHDESELPRALDKPYTVQFRLL 2309  
Db 2221 IYRTTLAQLPHNYDDPKSLRYPKXPIINTPVASISVHDESELPRALDKPYTVQFRLL 2280  
Qy 2310 ETEBRTKPICVFNHSHILVSGTGWMSARGCEVYFRNESHVSCCNHMTSPAVLMDVSRRE 2369  
Db 2281 ETEBRTKPICVFNHSHILVSGTGWMSARGCEVYFRNESHVSCCNHMTSPAVLMDVSRRE 2340  
Qy 2370 NGBILPLKLTLYVALGVTLAALLITPEFTLRLILSNQHIRRNLTAALGAOLVELLG 2429  
Db 2341 NGBILPLKLTLYVALGVTLAALLITPEFTLRLILSNQHIRRNLTAALGAOLVELLG 2400  
Qy 2430 INQADLPRACTVIALILHFTYLCITFSPALLBAHLTRALTEVDVNTGPMRFYTMGMV 2489  
Db 2401 INQADLPRACTVIALILHFTYLCITFSPALLBAHLTRALTEVDVNTGPMRFYTMGMV 2460  
Qy 2490 PAFITGLAVGLDDEGYNPDPCWLSIYDTLINSFAGVAVAFVMSVFLYTLARASCAQ 2549  
Db 2461 PAFITGLAVGLDDEGYNPDPCWLSIYDTLINSFAGVAVAFVMSVFLYTLARASCAQ 2520



QY 2550 RQFEKKGPVSGIQPSFAVLLLSATWLLLSVNSDTLLFHYLPATCNCTIQGPPIFLSY 2609  
DB 2521 RQFEKKGPVSGIQPSFAVLLLSATWLLLSVNSDTLLFHYLPATCNCTIQGPPIFLSY 2580  
QY 2610 VVLSKEVRKALKLACSRKPSDDPALTTKSTLTSYNCPSPYADGRLYOPYGDSSAGSLHST 2669  
DB 2581 VVLSKEVRKALKLACSRKPSDDPALTTKSTLTSYNCPSPYADGRLYOPYGDSSAGSLHST 2640  
QY 2670 SRSGKQPSYIPLRLREBSALNPGQPPGLGDPGSLFLBGOQDQDHPDSTDSDLSLEDD 2729  
DB 2641 SRSGKQPSYIPLRLREBSALNPGQPPGLGDPGSLFLBGOQDQDHPDSTDSDLSLEDD 2700  
QY 2730 QSGSYASTHSSDSEEBEEREEBAAPFEGCGMSLIGPGAFRIPLHSTPDGSGPGKAP 2789  
DB 2701 QSGSYASTHSSDSEEBEEREEBAAPFEGCGMSLIGPGAFRIPLHSTPDGSGPGKAP 2760  
QY 2790 WPGDFGTAAESSGNGAPBERLRENGDALREGLIGPLPSSSAQPHKGLIKKCLPTISE 2849  
DB 2761 WPGDFGTAAESSGNGAPBERLRENGDALREGLIGPLPSSSAQPHKGLIKKCLPTISE 2820  
QY 2850 KSSLILRLPLEQCTGSSRSGSSASGSRGPPRPQSLQEOQLNGWPIAMSIIKAGTVDE 2909  
DB 2821 KSSLILRLPLEQCTGSSRSGSSASGSRGPPRPQSLQEOQLNGWPIAMSIIKAGTVDE 2880  
QY 2910 DSSGSEFLPFNFLLH 2923  
DB 2881 DSSGSEFLPFNFLLH 2894

RESULT 5  
AAU07053  
ID AAU07053 standard; Protein; 2956 AA.  
AC AAU07053;  
DT 24-OCT-2001 (first entry)  
DE Human Flamingo polypeptide.  
XX Flamingo; human; G-protein coupled receptor; diabetes; protozoan; cancer;  
KW signal transduction pathway; bacterial; fungal; viral; anorexia; bulimia;  
KW asthma; Parkinson's disease; acute heart failure; HIV; obesity; anxiety;  
KW hypertension; hypertension; urinary retention; angina pectoris; allergy;  
KW myocardial infarction; stroke; ulcer; schizophrenia; psychotic disorder;  
KW benign prostatic hyperplasia; neurological disorder; manic depression;  
KW delirium; dementia; severe mental retardation; Huntington's disease;  
KW Gilles de la Tourette's syndrome; antibacterial; antifungal; antiviral;  
KW antiprotozoal; anti-HIV; anorectic; antianorectic; antiparkinsonian;  
KW cardiac; cerebroprotective; neuroprotective; antidepressant;  
KW anticonvulsant; antitense therapy; gene therapy.  
OS Homo sapiens.  
XX  
XX WO200161003-A1.  
PN 23-AUG-2001.  
XX  
XX 19-FEB-2001; 2001WO-GB00680.  
PF 19-FEB-2001; 2000GB-0004196.  
PR 19-FEB-2001; 2000GB-0004196.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX Testa TM;  
XX  
XX WPI; 2001-502792/55.  
DR N-PSDB; AAS11677.  
XX  
XX An isolated Flamingo polypeptide useful for treating diseases such as  
PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
PT osteoporosis -  
XX

PS Claim 2; Page 28-29; 66pp; English.  
XX  
XX The sequence represents a human Flamingo polypeptide. Flamingo is a  
CC member of the G-protein coupled receptor family, which is involved in  
CC signal transduction pathways. By screening to identify compounds that  
CC stimulate or inhibit the function or level of the protein, treatments can  
CC be developed for various diseases and bacterial, fungal, protozoan and  
CC viral infections, including HIV, cancer, diabetes, obesity, anorexia,  
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypertension,  
CC hypertension, urinary retention, angina pectoris, myocardial infarction,  
CC stroke, ulcers, allergies and benign prostatic hyperplasia. Also  
CC treatable are psychotic and neurological disorders such as anxiety,  
CC schizophrenia, manic depression, delirium, dementia, severe mental  
CC retardation, Huntington's disease and Gilles de la Tourette's syndrome.  
XX

SQ Sequence 2956 AA;

Query Match 97.1%; Score 2837; DB 22; Length 2956;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESPATGVLPPPPPLILLLLPPPLIGDQVPCRSIGSRGSSGACAPMGWLCPS 60  
DB 1 MESPATGVLPPPPPLILLLLPPPLIGDQVPCRSIGSRGSSGACAPMGWLCPS 60  
QY 61 SASNIMLYTSRCRDAGTELGHVPHHDGLRVWCPESEAHIPLPAPBECGPMSCRLGIG 120  
DB 61 SASNIMLYTSRCRDAGTELGHVPHHDGLRVWCPESEAHIPLPAPBECGPMSCRLGIG 120  
QY 121 GHLSPOGKLTPEBHPCLKAPRLRCQSCGLADAPGIRABERSPEBSLGRRRNNVTAQ 180  
DB 121 GHLSPOGKLTPEBHPCLKAPRLRCQSCGLADAPGIRABERSPEBSLGRRRNNVTAQ 180  
QY 181 FQPPSYOATVPENOPAGTFVASISRAIDPDEGAGRLLEYTMDALFDSRSNQFSLDPVTA 240  
DB 181 FQPPSYOATVPENOPAGTFVASISRAIDPDEGAGRLLEYTMDALFDSRSNQFSLDPVTA 240  
QY 241 VTTAEELDEIKSTVFRVTAQDHGMPPRSALATVLTITVTDNDHPVEQOEYKESLRE 300  
DB 241 VTTAEELDEIKSTVFRVTAQDHGMPPRSALATVLTITVTDNDHPVEQOEYKESLRE 300  
QY 301 NIEVGYVUTVATATGDAAPPNNILYRLLEGSGSPSEVFEIDPDSGVTRTGPVDRREV 360  
DB 301 NIEVGYVUTVATATGDAAPPNNILYRLLEGSGSGSPSEVFEIDPDSGVTRTGPVDRREV 360  
QY 361 ESYQUTVEASDQGRDPSPSTTAAVFLSVEDNDNAPQSEKRYVQVVEDVTPGAPVLR 420  
DB 361 ESYQUTVEASDQGRDPSPSTTAAVFLSVEDNDNAPQSEKRYVQVVEDVTPGAPVLR 420  
QY 421 VITASDRDKSNNAVHYSIMSGNARQFYLDAGTGLADVSPLDYETTKETTLRVAQDDG 480  
DB 421 VITASDRDKSNNAVHYSIMSGNARQFYLDAGTGLADVSPLDYETTKETTLRVAQDDG 480  
QY 481 RPLSNVSGLVTVQVLDINDNAPFVSTPPQATVLESPVGLTGLVQVLDINDAGNML 540  
DB 481 RPLSNVSGLVTVQVLDINDNAPFVSTPPQATVLESPVGLTGLVQVLDINDAGNML 540  
QY 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVEDYFSFGVARDGTPLATASASVTV 600  
DB 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVEDYFSFGVARDGTPLATASASVTV 600  
QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYQTSGNRRNFSITS 660  
DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDASHVITYQTSGNRRNFSITS 660  
QY 661 OSGGGLVSLALPLDYKLEROYVLAVTASDGRDPAQIVVNTDANTHRPVQSSHYTN 720  
DB 661 OSGGGLVSLALPLDYKLEROYVLAVTASDGRDPAQIVVNTDANTHRPVQSSHYTN 720  
QY 721 VNEDEPAGTIVVLSATDEDTGENARITYFMEDSIPQFRIDADTAVTTOAELDYEDQVS 780  
DB 721 VNEDEPAGTIVVLSATDEDTGENARITYFMEDSIPQFRIDADTAVTTOAELDYEDQVS 780



QY 781 YTLAIITARDNGI POKSDTTTLEIIVNDVNDAPQFLRDSYGSVYEDVPPTSVLQISAT 840  
Db 781 YTLAIITARDNGI POKSDTTTLEIIVNDVNDAPQFLRDSYGSVYEDVPPTSVLQISAT 840  
QY 841 DRDGLNGRVRVYTOGGDGDGDFIVESTSGIVARTLRDRENAQVYLAAYDKMPP 900  
Db 841 DRDGLNGRVRVYTOGGDGDGDFIVESTSGIVARTLRDRENAQVYLAAYDKMPP 900  
QY 901 ARTMEVTVTVLDVNDNPPVEODEFDVFEENSPIGLAVARTATPDEGTNAQIMYQI 960  
Db 901 ARTMEVTVTVLDVNDNPPVEODEFDVFEENSPIGLAVARTATPDEGTNAQIMYQI 960  
QY 961 VEGNIPEVFOQDIFSGEITLALVDLDYEDRPEYVLYIQATSAPLYSRATVHRLIDRNDP 1020  
Db 961 VEGNIPEVFOQDIFSGEITLALVDLDYEDRPEYVLYIQATSAPLYSRATVHRLIDRNDP 1020  
QY 1021 PVLENFELFNNTYNNSSSPFGALIGRPAMDIDISLTYSPERGNELSLVILNASTG 1080  
Db 1021 PVLENFELFNNTYNNSSSPFGALIGRPAMDIDISLTYSPERGNELSLVILNASTG 1080  
QY 1081 ELKLSRALDNNRPLEAIVSVSDGVSHTAOCALRVTLITDEMILTHSITLRLEDMSPER 1140  
Db 1081 ELKLSRALDNNRPLEAIVSVSDGVSHTAOCALRVTLITDEMILTHSITLRLEDMSPER 1140  
QY 1141 FLSPILGLFIOAVATATTPPDHVVVNVQRTDAPGGHILNVSLSVQPPGCGPPFL 1200  
Db 1141 FLSPILGLFIOAVATATTPPDHVVVNVQRTDAPGGHILNVSLSVQPPGCGPPFL 1200  
QY 1201 PSEDLORLYNRSLSLTAISAQVLPDPDNCIAREPCENYRCVSLRPFSSAFTIASS 1260  
Db 1201 PSEDLORLYNRSLSLTAISAQVLPDPDNCIAREPCENYRCVSLRPFSSAFTIASS 1260  
QY 1261 VLFPRIPHVGGRLRCRCPGFTGDCETEVDCYSRPGCGPHRCGRSGGTCLCRDGYT 1320  
Db 1261 VLFPRIPHVGGRLRCRCPGFTGDCETEVDCYSRPGCGPHRCGRSGGTCLCRDGYT 1320  
QY 1321 EHCVSARSRCCTGVCKNGCTCVNLVYGKCCPGSDPEKPYCOYTTSFPAHSITTF 1380  
Db 1321 EHCVSARSRCCTGVCKNGCTCVNLVYGKCCPGSDPEKPYCOYTTSFPAHSITTF 1380  
QY 1381 RGLRQRHFHTALSFATKRDGLLLYNGRFRBKHDPAVALRYIOQVOLTSSAGSITTVS 1440  
Db 1381 RGLRQRHFHTALSFATKRDGLLLYNGRFRBKHDPAVALRYIOQVOLTSSAGSITTVS 1440  
QY 1441 PVEGVSDGQWHTVOLKTYNKPILLGOTGLPQGSBQKVAVTVDGCOTGVALRFGSVLG 1500  
Db 1441 PVEGVSDGQWHTVOLKTYNKPILLGOTGLPQGSBQKVAVTVDGCOTGVALRFGSVLG 1500  
QY 1501 NYSQAAGTQGSKSLDLTGPIILLGVPLPESFPVRMRQFVGCNRNLQVDSRHIMAD 1560  
Db 1501 NYSQAAGTQGSKSLDLTGPIILLGVPLPESFPVRMRQFVGCNRNLQVDSRHIMAD 1560  
QY 1561 FIANGTVPGCPACKNYCDSVTGANGCTCVNQMDAFSCBCLPFGGSKCAQEMANPQFL 1620  
Db 1561 FIANGTVPGCPACKNYCDSVTGANGCTCVNQMDAFSCBCLPFGGSKCAQEMANPQFL 1620  
QY 1621 GSSIVAHGSLSPISQPMYLSIMFRTQADGVLLQAITRGRSTITTLQLRBGHVALSVET 1680  
Db 1621 GSSIVAHGSLSPISQPMYLSIMFRTQADGVLLQAITRGRSTITTLQLRBGHVALSVET 1680  
QY 1681 GLQASSRLRPERGRANDDMHQAOLALASGSGRAIISFDYGOQRAAGNIGPRLHGHLS 1740  
Db 1681 GLQASSRLRPERGRANDDMHQAOLALASGSGRAIISFDYGOQRAAGNIGPRLHGHLS 1740  
QY 1741 NITVGGIPGPAAGVARGFGLQGVRSVDPREGVNSLDPSSHGSLINTEGCSLPDPCSN 1800  
Db 1741 NITVGGIPGPAAGVARGFGLQGVRSVDPREGVNSLDPSSHGSLINTEGCSLPDPCSN 1800  
QY 1801 PCPANSYCSNDWDSYSCSDPGYTGDNCTNVCIDINPCESVCTRKPSAPHGTYCPCPN 1860  
Db 1801 PCPANSYCSNDWDSYSCSDPGYTGDNCTNVCIDINPCESVCTRKPSAPHGTYCPCPN 1860  
QY 1861 YLGPVCEFRIDQPCPRGMWHPITGCPNCVDVSKGFPDDCMKTSGECHCKENHVRPPSPPT 1920

Db 1861 YLGPVCEFRIDQPCPRGMWHPITGCPNCVDVSKGFPDDCMKTSGECHCKENHVRPPSPPT 1920  
QY 1921 CILCDYPPGSLRVCDDPEDGOCPCRPVYGRCCDCDNFPAVTNNGCVNDDSCPRAT 1980  
Db 1921 CILCDYPPGSLRVCDDPEDGOCPCRPVYGRCCDCDNFPAVTNNGCVNDDSCPRAT 1980  
QY 1981 EAGIWPRTFGLPAAPCPKSGFGTAVRCHDRHMLPMLFNCTSIYESLKGFAERL 2040  
Db 1981 EAGIWPRTFGLPAAPCPKSGFGTAVRCHDRHMLPMLFNCTSIYESLKGFAERL 2040  
QY 2041 QRNESGLDSRSQLALLRNATQHTAGIFGSVNVKAYQIATLALHESFORGEGSATQ 2100  
Db 2041 QRNESGLDSRSQLALLRNATQHTAGIFGSVNVKAYQIATLALHESFORGEGSATQ 2100  
QY 2101 DVHFTENLRYVGSALLDPTANKRHEMLIQOTEGTANLQHYEAYASALQNMHHTYLSF 2160  
Db 2101 DVHFTENLRYVGSALLDPTANKRHEMLIQOTEGTANLQHYEAYASALQNMHHTYLSF 2160  
QY 2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGEQPPDLFTTVILPESVPERTPPVVRPAG 2220  
Db 2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGEQPPDLFTTVILPESVPERTPPVVRPAG 2220  
QY 2221 PGSAQEPBEELARRQRHPELSQGEAVASVITYTLAQLPHNYDPDKRSIRVPEPIINT 2280  
Db 2221 PGSAQEPBEELARRQRHPELSQGEAVASVITYTLAQLPHNYDPDKRSIRVPEPIINT 2280  
QY 2281 PVVISVHDEBELLPALDKPVTVOQRLFETERTPICVFMHSHLVSQSGMSARGCE 2340  
Db 2281 PVVISVHDEBELLPALDKPVTVOQRLFETERTPICVFMHSHLVSQSGMSARGCE 2340  
QY 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPKTLTYVALGVTALALLTFPEFLT 2400  
Db 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPKTLTYVALGVTALALLTFPEFLT 2400  
QY 2401 LRILRNGGIRRNLTALAGLQVFLGIGNADLPACTVIALILHPLYLCTFSWALLR 2460  
Db 2401 LRILRNGGIRRNLTALAGLQVFLGIGNADLPACTVIALILHPLYLCTFSWALLR 2460  
QY 2461 ALHLRYALTEVRDVTNPMRFYMLGMGPAPITGLAVGLDPREGYNPFCMLSIYDTLI 2520  
Db 2461 ALHLRYALTEVRDVTNPMRFYMLGMGPAPITGLAVGLDPREGYNPFCMLSIYDTLI 2520  
QY 2521 WSPAGVAVASVFLYTLAARASCAQORQGEKKGPVSGLOPSFAVILLISATVLLAL 2580  
Db 2521 WSPAGVAVASVFLYTLAARASCAQORQGEKKGPVSGLOPSFAVILLISATVLLAL 2580  
QY 2581 LSVNSDTLLEHYLFATCNCIOGPFIFLSTYVLSKEVRKALKACSRKPSDPALTTKSTL 2640  
Db 2581 LSVNSDTLLEHYLFATCNCIOGPFIFLSTYVLSKEVRKALKACSRKPSDPALTTKSTL 2640  
QY 2641 TSSYNCPSYADRLTQPYGDSAGSLHSTRSGKSQPSYIFLLAREESALNPGQGPGLG 2700  
Db 2641 TSSYNCPSYADRLTQPYGDSAGSLHSTRSGKSQPSYIFLLAREESALNPGQGPGLG 2700  
QY 2701 DPGSLFLEQDQDHDPTDSDLSLEBDQSGSVASTHSSDSEBEEBEEBAAPFGEQ 2760  
Db 2701 DPGSLFLEQDQDHDPTDSDLSLEBDQSGSVASTHSSDSEBEEBEEBAAPFGEQ 2760  
QY 2761 WDSLDPGAEURLPHSTPDGPGPKAPWPDPGTAAESSNGAPBXRILRNGALSR 2820  
Db 2761 WDSLDPGAEURLPHSTPDGPGPKAPWPDPGTAAESSNGAPBXRILRNGALSR 2820  
QY 2821 EGSIGLPGSSAQPHKG 2837  
Db 2821 EGSIGLPGSSAQPHKG 2837  
RESULT 6  
AAU74826  
ID AAU74826 standard; protein; 2936 AA.  
XX  
AC AAU74826;

XX 23-APR-2002 (first entry)  
 XX Human REPTR 9 protein.  
 DE  
 XX  
 KM REPTR: human; antiinflammatory; cytostatic; immunosuppressive;  
 KM antiviral; anti-HIV; antitumor; anticonvulsant; nootropic;  
 KM neuroprotective; anti-allergic; antibody; immunogen; endometriosis;  
 KM gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KM Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KM endocrine disorder; hypothyroidism; Kallman's disease;  
 KM autoimmune disease; inflammatory disease; infertility; receptor;  
 KM acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KM allergic; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KM systemic lupus erythematosus; cell proliferative disorder;  
 KM cancer; developmental disorder; Duchenne muscular dystrophy;  
 KM Becker muscular dystrophy; neurological disorder; epilepsy;  
 KM Alzheimer's disease; Huntington's disease; reproductive disorder.  
 XX  
 OS Homo sapiens.  
 PN MO200198354-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001MO-US19942.  
 XX  
 PR 21-JUN-2000; 2000US-214027P. — *Chick*  
 PR 25-JUN-2000; 2000US-228045P.  
 PR 12-DEC-2000; 2000US-255104P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Griffen JA, Kalliock DA, Tribouley CM, Yue H, Nguyen DB, Tang YT;  
 PI Lal P, Policky JL, Azimzai Y, Lu DM, Grail R, Yao MG, Burford N;  
 PI Hafalja AJ, Baughn MR, Bandman O, Patterson C, Yang J, Xu Y;  
 PI Gandhi AR, Warren BA, Ding L, Sanjanwala NS, Duggan BM, Lu Y;  
 XX WPI; 2002-090432/12.  
 DR N-PSDB; ABK15177.  
 XX  
 PT Twelve human receptors (referred to as REPTR-1 to REPTR-12), useful in  
 PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
 PT proliferative (e.g. cancer) disorders -  
 XX  
 PS Claim 53; Page 131-138; 157pp; English.  
 XX  
 CC This invention relates to twelve human receptors cDNA sequences  
 CC referred to as REPTR-1 to REPTR-12), and the proteins encoded thereby.  
 CC The proteins of the invention may have antiinflammatory, cytostatic,  
 CC immunosuppressive, antiviral, anti-HIV, antitumor, muscular active  
 CC general, anticonvulsant, nootropic, neuroprotective, antiallergic  
 CC activities. The sequences of the invention may be used to produce REPTR  
 CC agonists or antagonists, and the protein sequences may be used to raise  
 CC anti-REPTR antibodies. These molecules and the REPTR polynucleotides and  
 CC polypeptides of the invention are useful in the diagnosis, treatment and  
 CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,  
 CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),  
 CC endocrine (e.g. hypothyroidism disorder, Kallman's disease), autoimmune/  
 CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS)),  
 CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,  
 CC multiple sclerosis, systemic lupus erythematosus), cell proliferative  
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
 CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,  
 CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)  
 CC disorders. Numerous other examples of each disorder are given in the  
 CC specification. The present sequence represents the human REPTR9 protein  
 CC sequence of the invention.  
 XX  
 XX Sequence 2936 AA;  
 Query Match 67.0%; Score 1958; DB 23; Length 2936;

Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1958; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 PPPPLLLLLLLLLLPPPLLDGVGPGCSLGRSGSSGACAPMCMCLPPSSASNTMTYTSNC 72  
 DB 14 PPPPLLLLLLLLLLPPPLLDGVGPGCSLGRSGSSGACAPMCMCLPPSSASNTMTYTSNC 73  
 QY 73 RDAGELTGLVPHHGLRWMCPESEBAHIPLPAPAGCPCWSCGLGIGHLSPQGLTLP 132  
 DB 74 RDAGELTGLVPHHGLRWMCPESEBAHIPLPAPAGCPCWSCGLGIGHLSPQGLTLP 133  
 QY 133 EHPHCKAPRLRCQSCCKLAQAPGLRAGERSPEESLGGRRKRVNTAPQCPSPSYQATVPE 192  
 DB 134 EHPHCKAPRLRCQSCCKLAQAPGLRAGERSPEESLGGRRKRVNTAPQCPSPSYQATVPE 193  
 QY 193 NOPAGPVASLAIIDDEGRARLEYTMALPDSRNOFPISLDPVGAATTAEELEDRK 252  
 DB 194 NOPAGPVASLAIIDDEGRARLEYTMALPDSRNOFPISLDPVGAATTAEELEDRK 253  
 QY 253 STHVFRTAODHGMPPRSALATLTLLVTDNDHPVFEQOEYKESLREMLEVGYELTVR 312  
 DB 254 STHVFRTAODHGMPPRSALATLTLLVTDNDHPVFEQOEYKESLREMLEVGYELTVR 313  
 QY 313 ATDGAAPVNAITLRLBGSQSPSEVFEBIDPSGYITRGPVDRREVESTYQTVASDQ 372  
 DB 314 ATDGAAPVNAITLRLBGSQSPSEVFEBIDPSGYITRGPVDRREVESTYQTVASDQ 373  
 QY 373 GRDGRSTTAAPVLSVEDNDNNAPOFSEKRVYVQREBVTGAPVLTATASDRDGSNA 432  
 DB 374 GRDGRSTTAAPVLSVEDNDNNAPOFSEKRVYVQREBVTGAPVLTATASDRDGSNA 433  
 QY 433 VVHYSIMSGNARGQFLDAGTALDVSPLDYETTKETTLRYAODGGRPPLSNVGLVT 492  
 DB 434 VVHYSIMSGNARGQFLDAGTALDVSPLDYETTKETTLRYAODGGRPPLSNVGLVT 493  
 QY 493 VQVLDINDNAPLFPVSTPPQATVLESYPLGVLVAVQALDADAGDNARLEYLAGVGHDP 552  
 DB 494 VQVLDINDNAPLFPVSTPPQATVLESYPLGVLVAVQALDADAGDNARLEYLAGVGHDP 553  
 QY 553 FTINNGTMIIVAAELDREEDVPYSPGEARHGPALFASASVSTVLDVNDNNTFPQ 612  
 DB 554 FTINNGTMIIVAAELDREEDVPYSPGEARHGPALFASASVSTVLDVNDNNTFPQ 613  
 QY 613 PEYVALNEDAAVGVSVTVASVDRDAHSVITYQTISGTRNRFPSTISQSGGLVSLAP 672  
 DB 614 PEYVALNEDAAVGVSVTVASVDRDAHSVITYQTISGTRNRFPSTISQSGGLVSLAP 673  
 QY 673 LDYKERQVYLAVTASDGTROTQALVNVNTDANTHRPVFGSHYTVVNEBRPAGTIV 732  
 DB 674 LDYKERQVYLAVTASDGTROTQALVNVNTDANTHRPVFGSHYTVVNEBRPAGTIV 733  
 QY 733 LISATDEDEGENARTTYFEDSIPOFRIDADGAVTQALDYEDOVSTTLATITANDGI 792  
 DB 734 LISATDEDEGENARTTYFEDSIPOFRIDADGAVTQALDYEDOVSTTLATITANDGI 793  
 QY 793 POKSDTYLEILLVNDVNDNAPQLRDSYQGSYVEDPPTSYQLSATDRDGLNGRVY 852  
 DB 794 POKSDTYLEILLVNDVNDNAPQLRDSYQGSYVEDPPTSYQLSATDRDGLNGRVY 853  
 QY 853 TPOGGDDGCGDITVESTGIVTTLRLDRBNAAQVYLAAYVDKMPARTPMEVTVTL 912  
 DB 854 TPOGGDDGCGDITVESTGIVTTLRLDRBNAAQVYLAAYVDKMPARTPMEVTVTL 913  
 QY 913 DVNDNPPVEQGEFVDFVENSPIGLAVARVATPDBETNAQIMQIYEGNIPEVFQD 972  
 DB 914 DVNDNPPVEQGEFVDFVENSPIGLAVARVATPDBETNAQIMQIYEGNIPEVFQD 973  
 QY 973 IFSGELTALVDLDYDRPEYVLIQATSAPLVSRATVHRLDRDNDPVLGNFELFNN 1032  
 DB 974 IFSGELTALVDLDYDRPEYVLIQATSAPLVSRATVHRLDRDNDPVLGNFELFNN 1033  
 QY 1033 YVTNRSSSPGCAIGRVPAHNDPDISLTYFERGENELSVLIANSTGLKLSRALDNR 1092

Db 1034 YVNRSSSPGCAIGRVAHPADPDISDLSITFSERGENELSLVLNASTGBELKLSRALDNMR 1093  
 QY PLEAIMSVLSDGVSVTAOCALRTVITITDEMLTHSITLRLSDMSPERFLSPILGLFIQA 1152  
 Db 1094 PLEAIMSVLSDGVSVTAOCALRTVITITDEMLTHSITLRLSDMSPERFLSPILGLFIQA 1153  
 QY 1153 VAATLATPPDHVVVFNVRDTPAPGHIILNLSVGVQPPGPGQPPFLPSEDLQERLYLN 1212  
 Db 1154 VAATLATPPDHVVVFNVRDTPAPGHIILNLSVGVQPPGPGQPPFLPSEDLQERLYLN 1213  
 QY 1213 RSLTALTAQVRLPPDDNICLRPEPCENMRCSVSLRFPSSAPFLASSSVLRPIHPVGL 1272  
 Db 1214 RSLTALTAQVRLPPDDNICLRPEPCENMRCSVSLRFPSSAPFLASSSVLRPIHPVGL 1273  
 QY 1273 RCRCPGFTGYCEFEVDLCTSRPCGPRGRCSRSRGVTCRCRDYTSERHCEVSARSRC 1332  
 Db 1274 RCRCPGFTGYCEFEVDLCTSRPCGPRGRCSRSRGVTCRCRDYTSERHCEVSARSRC 1333  
 QY 1333 TPGVCKNGGTCVNLIVGSPKDCPSGDFEKPQVTTTSPPAHSFTFRGLRQRFHTLA 1392  
 Db 1334 TPGVCKNGGTCVNLIVGSPKDCPSGDFEKPQVTTTSPPAHSFTFRGLRQRFHTLA 1393  
 QY 1393 LSPATKEDGILLNNGRPNRKNDFALAVIOBQVLTSSAGESTTTPSPVPGVSDGQW 1452  
 Db 1394 LSPATKEDGILLNNGRPNRKNDFALAVIOBQVLTSSAGESTTTPSPVPGVSDGQW 1453  
 QY 1453 HTVOLKTYNKEPLAGTGLPGSPSEBOKVAVVTVVDCDGTVALRFGSVLGNYSCAAQGTGG 1512  
 Db 1454 HTVOLKTYNKEPLAGTGLPGSPSEBOKVAVVTVVDCDGTVALRFGSVLGNYSCAAQGTGG 1513  
 QY 1513 SKKSLDTLGPILLGVPDLPSFPVNRKQFVGCNRNLTQVDSRHIMADPIANNGTVPQCP 1572  
 Db 1514 SKKSLDTLGPILLGVPDLPSFPVNRKQFVGCNRNLTQVDSRHIMADPIANNGTVPQCP 1573  
 QY 1573 AKONVCSNTNGNCTCNOWDABSCCEPLRGGSCQOEBANPQHPFGSSLVAMHGSL 1632  
 Db 1574 AKONVCSNTNGNCTCNOWDABSCCEPLRGGSCQOEBANPQHPFGSSLVAMHGSL 1633  
 QY 1633 PISOPWYLSLFRTRQADGVLLQAITRGRSTTTLQREGHWLSEVGTGLQASSLRLBPG 1692  
 Db 1634 PISOPWYLSLFRTRQADGVLLQAITRGRSTTTLQREGHWLSEVGTGLQASSLRLBPG 1693  
 QY 1693 RANDGDMHAOLALGASGPGHAILSPYGGQRAAGNTGPRHGLHLSNTVGGIPGPAG 1752  
 Db 1694 RANDGDMHAOLALGASGPGHAILSPYGGQRAAGNTGPRHGLHLSNTVGGIPGPAG 1753  
 QY 1753 GVANGFRGCTGGVRVSDTPEGVNSLDPBHGSIINVEGCSLPDPCDSNPPCANSTCSNDW 1812  
 Db 1754 GVANGFRGCTGGVRVSDTPEGVNSLDPBHGSIINVEGCSLPDPCDSNPPCANSTCSNDW 1813  
 QY 1813 DSYSCSCDPGYGDNCTVNCDLNPECHOSVCTRKPSAPHGTYCECPVNYLGPYCETRIQ 1872  
 Db 1814 DSYSCSCDPGYGDNCTVNCDLNPECHOSVCTRKPSAPHGTYCECPVNYLGPYCETRIQ 1873  
 QY 1873 PCPRGMWHPPTGCPNCNCSVSKGFPDCKNTSGECHKENHTRPPSPCTLLCDCTPTGSL 1932  
 Db 1874 PCPRGMWHPPTGCPNCNCSVSKGFPDCKNTSGECHKENHTRPPSPCTLLCDCTPTGSL 1933  
 QY 1933 SRVCDPEHGOCPCKRGVIGROCDRCNDNPAEYTNNGCE 1970  
 Db 1934 SRVCDPEHGOCPCKRGVIGROCDRCNDNPAEYTNNGCE 1971  
 RESULT 7  
 ABB11404 ID ABB11404 standard; peptide; 2560 AA.  
 AC ABB11404;  
 XX 11-JAN-2002 (first entry)  
 XX Human FLAMINGO 1 homologue, SEQ ID NO:11774.  
 XX

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; acthna; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiaesthetic; antiarthritis; haemostatic; antiatherosclerotic;  
 KW cytosolic; osteopathic; vasodilator; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antitumor.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200157188-A2.  
 PN  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001, 2001WO-US03800.  
 XX  
 PR 03-FEB-2000, 2000US-0496914.  
 PR 27-APR-2000, 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457740/49.  
 DR N-PSDB; ABA08648.  
 XX  
 PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 PS Claim 20; Page 179-181; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug

CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX Sequence 2560 AA;

Query Match 65.6%; Score 1917; DB 22; Length 2560;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2317; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 380 STTAATFLSLEDNDNAPQSEKRYVQVREDYTPGAPVLRATASDRDGSNAVHYXSM 439  
 Db 13 STTAATFLSLEDNDNAPQSEKRYVQVREDYTPGAPVLRATASDRDGSNAVHYXSM 72  
 QY 440 SGNARQFTLDAQTALDVNSPLDYETKXETLVRADQGRPLSNVSGLVTVQVLIN 499  
 Db 73 SGNARQFTLDAQTALDVNSPLDYETKXETLVRADQGRPLSNVSGLVTVQVLIN 132  
 QY 500 DNAPFVSTPQATVLESVPLGLVLAHQALIDADADNARLEYRLAGVGHDPPTINNQT 559  
 Db 133 DNAPFVSTPQATVLESVPLGLVLAHQALIDADADNARLEYRLAGVGHDPPTINNQT 192  
 QY 560 GNISVAELDRREVDYTSFGVEARDHGTPTLTASASVTVLDVNDNPTFQPEYVRL 619  
 Db 193 GNISVAELDRREVDYTSFGVEARDHGTPTLTASASVTVLDVNDNPTFQPEYVRL 252  
 QY 620 NEDAAVGVSVTVASAVDRDAHVVITYQITSGNTRNFSITSGGGGLVSLALPLDYKLR 679  
 Db 253 NEDAAVGVSVTVASAVDRDAHVVITYQITSGNTRNFSITSGGGGLVSLALPLDYKLR 312  
 QY 680 QVLAATASDGTRODPAQIVNNTDANTHRPVPQSHYTVNVEDPACTTVVLSATDE 739  
 Db 313 QVLAATASDGTRODPAQIVNNTDANTHRPVPQSHYTVNVEDPACTTVVLSATDE 372  
 QY 740 DTGENARITYFMEDSIPQFRIDADTGAVTQAELEDQVSYTLATARDNGIPQKSDYT 799  
 Db 373 DTGENARITYFMEDSIPQFRIDADTGAVTQAELEDQVSYTLATARDNGIPQKSDYT 432  
 QY 800 YLEILVNDVNDNAPQFLRDSYQSGVYEDVPPTSTVLQISATDRDSGLNCRVYTFQGGD 859  
 Db 433 YLEILVNDVNDNAPQFLRDSYQSGVYEDVPPTSTVLQISATDRDSGLNCRVYTFQGGD 492  
 QY 860 GGGDFIVESTSGIVRTLRRLDRNVNAQVTLRAVNDKMPARTPMEVTVTVLDVNDNP 919  
 Db 493 GGGDFIVESTSGIVRTLRRLDRNVNAQVTLRAVNDKMPARTPMEVTVTVLDVNDNP 552  
 QY 920 VEEODEFDVFEVENSPIGLAVARVATDDEGTNAQIMQIVEGNIPFYOLDIFSGELT 979  
 Db 553 VEEODEFDVFEVENSPIGLAVARVATDDEGTNAQIMQIVEGNIPFYOLDIFSGELT 612  
 QY 980 ALVDLDYEDRPVYVIAQTSAPLVSRATVHVRLLDRNDNPVLGNFEILLNNYVTRSS 1039  
 Db 613 ALVDLDYEDRPVYVIAQTSAPLVSRATVHVRLLDRNDNPVLGNFEILLNNYVTRSS 672  
 QY 1040 SPFGAIGVPAHDDPISLSTYSPERGNELSVLVLNASTGELKSRALDNNRPLEAINS 1099  
 Db 673 SPFGAIGVPAHDDPISLSTYSPERGNELSVLVLNASTGELKSRALDNNRPLEAINS 732  
 QY 1100 VLASDVHVSVAQALRVTIITDEMILTHSITLLEDMSPERFLSPLGLFQAVATLAT 1159  
 Db 733 VLASDVHVSVAQALRVTIITDEMILTHSITLLEDMSPERFLSPLGLFQAVATLAT 792  
 QY 1160 PPDHVVVENVQDTPAGGHIILNLSISVQPPGPGGPPFLPSEDLQERLYNRSILTAI 1219  
 Db 793 PPDHVVVENVQDTPAGGHIILNLSISVQPPGPGGPPFLPSEDLQERLYNRSILTAI 852  
 QY 1220 SAQRVLPEDDNI CLRPECNVRCVSLAFDSSAPRTASSVLPFRITHVVGGLRCRCPPG 1279  
 Db 853 SAQRVLPEDDNI CLRPECNVRCVSLAFDSSAPRTASSVLPFRITHVVGGLRCRCPPG 912  
 QY 1280 FTGDVCEFEVDLCYSRPGCRSREGGYTCLCDGYTGHCCEVSARSGRCTGVCKN 1339  
 Db 913 FTGDVCEFEVDLCYSRPGCRSREGGYTCLCDGYTGHCCEVSARSGRCTGVCKN 972

QY 1340 GGTCTVNLVGGFKCDPCPSGDFEKPYCOVTRSPPAHSIFTRGLRQRFHTLALSPATKE 1399  
 Db 973 GGTCTVNLVGGFKCDPCPSGDFEKPYCOVTRSPPAHSIFTRGLRQRFHTLALSPATKE 1032  
 QY 1400 RDGILLNKRPNKXIDPVALETIOBOVOLTEBAGSSTTVSPFPGVSDGQHTVQLKI 1459  
 Db 1033 RDGILLNKRPNKXIDPVALETIOBOVOLTEBAGSSTTVSPFPGVSDGQHTVQLKI 1092  
 QY 1460 YNKPILLGQTLQSGSEOKAVAVTVDCDTPVALPFGSVLGNYSCAQGTQGGSKSLDL 1519  
 Db 1093 YNKPILLGQTLQSGSEOKAVAVTVDCDTPVALPFGSVLGNYSCAQGTQGGSKSLDL 1152  
 QY 1520 TGPILLGVPDLPSFPVARMQPVGCMNTQVDSRHIDMADRIANNITVPGCCPAKKNVCD 1579  
 Db 1153 TGPILLGVPDLPSFPVARMQPVGCMNTQVDSRHIDMADRIANNITVPGCCPAKKNVCD 1212  
 QY 1580 SNTCHNGGTCVNMQWAFSGCEPLGFGKSCAQEMANPQHLGLSSLVAMHGLSLPISQPMY 1639  
 Db 1213 SNTCHNGGTCVNMQWAFSGCEPLGFGKSCAQEMANPQHLGLSSLVAMHGLSLPISQPMY 1272  
 QY 1640 LSLMFRTOADGVLLQAITRGRSTITLQRBGVNLSYEGTGLQASSLRBERGRANDGM 1699  
 Db 1273 LSLMFRTOADGVLLQAITRGRSTITLQRBGVNLSYEGTGLQASSLRBERGRANDGM 1332  
 QY 1700 HHAQALGASGGPQGAHILSFDYGOQRAEAGLPRLLHGLHSNITVGGIPGPAQVAKGR 1759  
 Db 1333 HHAQALGASGGPQGAHILSFDYGOQRAEAGLPRLLHGLHSNITVGGIPGPAQVAKGR 1392  
 QY 1760 GGLQGVRSIDTPREGVNSLDPHSGESINVEQCSLPDPCDSNCPANSTCSNDWDSYSGC 1819  
 Db 1393 GGLQGVRSIDTPREGVNSLDPHSGESINVEQCSLPDPCDSNCPANSTCSNDWDSYSGC 1452  
 QY 1820 DPGYGDNCTVNCDLNPNCEHOSVCTRKPSAPHGVCCECPNVLGYCETRIDIQPCPRGM 1879  
 Db 1453 DPGYGDNCTVNCDLNPNCEHOSVCTRKPSAPHGVCCECPNVLGYCETRIDIQPCPRGM 1512  
 QY 1880 GHPTGPNCDVSKGPDPCNKTSGBCHCKENHYRPGSPCTCLDCYPTGSLSRVCDPE 1939  
 Db 1513 GHPTGPNCDVSKGPDPCNKTSGBCHCKENHYRPGSPCTCLDCYPTGSLSRVCDPE 1572  
 QY 1940 DGQPCCKRGVIGROCDRCNDPFAVYTNCGEVNYSCEPRAIEAGTWMPRTRGLEPAAAPC 1999  
 Db 1573 DGQPCCKRGVIGROCDRCNDPFAVYTNCGEVNYSCEPRAIEAGTWMPRTRGLEPAAAPC 1632  
 QY 2000 PKGSFGTAVRCHDERGMPLPMLFNCTSIITSELKGPABERLQORNSGJDSGSOQLALL 2059  
 Db 1633 PKGSFGTAVRCHDERGMPLPMLFNCTSIITSELKGPABERLQORNSGJDSGSOQLALL 1692  
 QY 2060 RNATQHTAGYFGSDVKAAYOLATRLLAHESYORGFGLSATODVHFTENILRVGSLALDTA 2119  
 Db 1693 RNATQHTAGYFGSDVKAAYOLATRLLAHESYORGFGLSATODVHFTENILRVGSLALDTA 1752  
 QY 2120 NKRHWELLQOTBEGTAMLLQHTYEAASLAAQNMHTTISPTTIYTPNIVISVRLDKNF 2179  
 Db 1753 NKRHWELLQOTBEGTAMLLQHTYEAASLAAQNMHTTISPTTIYTPNIVISVRLDKNF 1812  
 QY 2180 AGAKLPRYBALRGSDPDLSTTVLLPESVFPETPPVAPAGGEAQOBEELARORRDE 2239  
 Db 1813 AGAKLPRYBALRGSDPDLSTTVLLPESVFPETPPVAPAGGEAQOBEELARORRDE 1872  
 QY 2240 LSGEAVASVIIYRTLAGLPHNYDPDRSLRVPKRP1INTPVVSIYVHDEBELPRALD 2299  
 Db 1873 LSGEAVASVIIYRTLAGLPHNYDPDRSLRVPKRP1INTPVVSIYVHDEBELPRALD 1932  
 QY 2300 KPVTVQFLLTERTERTKICVPMNHSILVSGTGGMSAGCEVFPNESHVSQCNHMTSF 2359  
 Db 1933 KPVTVQFLLTERTERTKICVPMNHSILVSGTGGMSAGCEVFPNESHVSQCNHMTSF 1992  
 QY 2360 AVLMDVSRRENGEILLPLKTLTYVALGVTLAALLTFPFLTLRLIRSNQHGIRNLTAL 2419  
 Db 1993 AVLMDVSRRENGEILLPLKTLTYVALGVTLAALLTFPFLTLRLIRSNQHGIRNLTAL 2052  
 QY 2420 GLAQVLELGINQADLPACTVIAILHFLYLCFSPWALLRALHLRYALTEVRDVTGPM 2479

Db 2053 GLAQLVLLGINQADLPFACVLAIIHLHFIYLCFFSWALLBALFLYALTEVDVMTGPM 2112  
 Qy 2480 RFFYMLGMGPALFTGLAVGLDPBGVGNPDRCWMLSTYDTLIMSAGVAVAVNSVLYI 2539  
 Db 2113 RFTYMLGMGPALFTGLAVGLDPBGVGNPDRCWMLSTYDTLIMSAGVAVAVNSVLYI 2172  
 Qy 2540 LAARASCAAROGKEKKGPVSGLOPSFAVLLLSATWLLLSVNSDTLLFHYLFATCNC 2599  
 Db 2173 LAARASCAAROGKEKKGPVSGLOPSFAVLLLSATWLLLSVNSDTLLFHYLFATCNC 2232  
 Qy 2600 IQGPFITLSYVLSKEYRKALKACSRKSPDPALTTKSTLTSYNSCPSPYADRLTYQPY 2659  
 Db 2233 IQGPFITLSYVLSKEYRKALKACSRKSPDPALTTKSTLTSYNSCPSPYADRLTYQPY 2292  
 Qy 2660 GDSAGSLHSTRSGSKSPSYIPPLIRESSALNPQGGPGIG 2700  
 Db 2293 GDSAGSLHSTRSGSKSPSYIPPLIRESSALNPQGGPGIG 2333

## RESULT 8

AAB42192  
 ID AAB42192. standard; Protein: 2405 AA.

AC AAB42192;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF1956 polypeptide sequence SEQ ID NO:3912.

Xx Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 Km vulnery; antiparasitic; antiparkinsonian; noctropic; neuroprotective;  
 Km anticovulsant; osteoporotic; antichratic; immunosuppressant; candidant;  
 Km immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 Km hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 Km antiviral; antibacterial; antifungal; antineumatic; antichyroid;  
 Km antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 Km neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 Km cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 Km cholesterol ester storage; systemic lupus erythematosus; infection;  
 Km severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 Km allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 Km bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 Km thrombosis; contraceptive.

OS Homo sapiens.

Xx MO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2000-602362/57.

XX N-PSDB; AAC76401.

PT Novel nucleic acids and peptides derived from open reading frame X,

CC useful for treating e.g. cancers, proliferative disorders,

CC neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 3067-3072; 5507p; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiparasitic; antiparkinsonian; noctropic; neuroprotective;  
 CC osteoporotic; anticovulsant; antichratic; immunosuppressant;  
 CC immunostimulant; candidant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antichyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 2405 AA.

Query Match 55.5%; Score 1623; DB 21; Length 2405;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1301 GRCRSREGYTCLCRDGYTGEHCEVARSGRCTPGVCKNGTCVNLVGGFKDCPSGPF 1360  
 Db 783 GRCRSREGYTCLCRDGYTGEHCEVARSGRCTPGVCKNGTCVNLVGGFKDCPSGPF 842  
 Qy 1361 EKPVCQVTRSPFPAHSFIFRGLRQRFHTLASFAFKERDGLLYNGRPNKDFVALE 1420  
 Db 843 EKPVCQVTRSPFPAHSFIFRGLRQRFHTLASFAFKERDGLLYNGRPNKDFVALE 902  
 Qy 1421 VIOBOVQLTFSAGESTTYSPPVPGVSDGQWHTVQLKYNRPLDQTLPGQSSQKVA 962  
 Db 903 VIOBOVQLTFSAGESTTYSPPVPGVSDGQWHTVQLKYNRPLDQTLPGQSSQKVA 962  
 Qy 1481 VVTVDCDTPVALRFSVLYGNSCAAGTQGSKSLDTLGPLLGGVDPDPSPFVRR 1540  
 Db 963 VVTVDCDTPVALRFSVLYGNSCAAGTQGSKSLDTLGPLLGGVDPDPSPFVRR 1022  
 Qy 1541 QFVGCNRNLOVDSRHIDMADFIANNGTVPCCPKXNVCSNTCHNGTCVNMDFSCC 1600  
 Db 1023 QFVGCNRNLOVDSRHIDMADFIANNGTVPCCPKXNVCSNTCHNGTCVNMDFSCC 1082  
 Qy 1601 PLRFGKSCAQEMANPOHPLGSLVAMHGLSLPIGQWYLSIMFRRODGVLLQAITTG 1660  
 Db 1083 PLRFGKSCAQEMANPOHPLGSLVAMHGLSLPIGQWYLSIMFRRODGVLLQAITTG 1142  
 Qy 1661 RSTTTQLRSGHVLSEGTGLQASSLRLEPGRANDGWHHAQLAGSGPGHAILISFD 1720  
 Db 1143 RSTTTQLRSGHVLSEGTGLQASSLRLEPGRANDGWHHAQLAGSGPGHAILISFD 1202  
 Qy 1721 YGQORABGNIGPRLHGLHLSNITVGGI PGPAGVARGFGGCIQVAVSDTPBGVNSLDES 1780  
 Db 1203 YGQORABGNIGPRLHGLHLSNITVGGI PGPAGVARGFGGCIQVAVSDTPBGVNSLDES 1262  
 Qy 1781 HGRSINVBQCSLPDCCDNPCCPANSYCSNDMYSQCDPGYGNCTNVCDLNCEHQ 1840  
 Db 1263 HGRSINVBQCSLPDCCDNPCCPANSYCSNDMYSQCDPGYGNCTNVCDLNCEHQ 1322  
 Qy 1841 SVCTRKPSAPHGTCECPRYTLAGPYCETRIDPCPRGMWGHPTCGPCNCDVSKGPPDCN 1900  
 Db 1323 SVCTRKPSAPHGTCECPRYTLAGPYCETRIDPCPRGMWGHPTCGPCNCDVSKGPPDCN 1382  
 Qy 1901 KTSGBCHKENHRRPFGSPICLLCDYCPYGSLSRVCDPBGQCPCKPGVYGRQCDRCNDP 1960  
 Db 1383 KTSGBCHKENHRRPFGSPICLLCDYCPYGSLSRVCDPBGQCPCKPGVYGRQCDRCNDP 1442  
 Qy 1961 FAEVTNGCEVNVDSGCPRAIEAGIWWPRTFRGLPAAPCPKSGFGTAVRHCDHRGMLP 2020  
 Db 1443 FAEVTNGCEVNVDSGCPRAIEAGIWWPRTFRGLPAAPCPKSGFGTAVRHCDHRGMLP 1502

QY 2021 NLFNCTSTTPESELKPAERLQNESGLDSGRSQQLALLRNATQHTAGYFGSDVAVAYQL 2080  
 DB 1503 NLFNCTSTTPESELKPAERLQNESGLDSGRSQQLALLRNATQHTAGYFGSDVAVAYQL 1562  
 QY 2081 ATRLIAHSTORGPGLSATQDVHFTENILLRVSALLDTANKHMLIQTEGTWMLLOH 2140  
 DB 1563 ATRLIAHSTORGPGLSATQDVHFTENILLRVSALLDTANKHMLIQTEGTWMLLOH 1622  
 QY 2141 YEAVASALAQNMRHTYLSPTITVTENIVISVRLDKGNFAGAKLPRYEALRGEOPDLET 2200  
 DB 1663 YEAVASALAQNMRHTYLSPTITVTENIVISVRLDKGNFAGAKLPRYEALRGEOPDLET 1682  
 QY 2201 TVILPESVFRETPPVVRPAPGEGAOPEBELARORRHPELSQGBAVASYIYRITLAGLLP 2260  
 DB 1683 TVILPESVFRETPPVVRPAPGEGAOPEBELARORRHPELSQGBAVASYIYRITLAGLLP 1742  
 QY 2261 HNYDDKXSLRVPKRPINTPVPVSTSYHDEDELLPRALDKPTVVOFRLLETERKPCY 2320  
 DB 1743 HNYDDKXSLRVPKRPINTPVPVSTSYHDEDELLPRALDKPTVVOFRLLETERKPCY 1802  
 QY 2321 FMNHSILVSGTGMGARGCEVVRNESHVSQCNMTSPAVLMDVSRRENGBILPLKTLT 2380  
 DB 1803 FMNHSILVSGTGMGARGCEVVRNESHVSQCNMTSPAVLMDVSRRENGBILPLKTLT 1862  
 QY 2381 YVALGVTLAALLTFPFLTLRLRSNOGIRRNITAAALGLAQVFLGLINQADLPFAC 2440  
 DB 1863 YVALGVTLAALLTFPFLTLRLRSNOGIRRNITAAALGLAQVFLGLINQADLPFAC 1922  
 QY 2441 VIATILHFLYCTFSWALLBALHLYRALTEVADVNTGMPRFYMLGNGVPATITGLANG 2500  
 DB 1923 VIATILHFLYCTFSWALLBALHLYRALTEVADVNTGMPRFYMLGNGVPATITGLANG 1982  
 QY 2501 DPEGYNDFPCWLSIYDTLWISFAPVAVASMSVFLYTLAARASCAAROGFEKGPVS 2560  
 DB 1983 DPEGYNDFPCWLSIYDTLWISFAPVAVASMSVFLYTLAARASCAAROGFEKGPVS 2042  
 QY 2561 GLQPSFAVLLLSATWLLALISVNSDTLLFHYLFATCNCIOGPFIFLSYVVLSEKVRAL 2620  
 DB 2043 GLQPSFAVLLLSATWLLALISVNSDTLLFHYLFATCNCIOGPFIFLSYVVLSEKVRAL 2102  
 QY 2621 KLACGRKSPDPALTTKSTLTVSSVNCSPYADGRLYQYVGSAGSLHSTSRGKQPSYI 2680  
 DB 2103 KLACGRKSPDPALTTKSTLTVSSVNCSPYADGRLYQYVGSAGSLHSTSRGKQPSYI 2162  
 QY 2681 PFLAREBSALNPGQPPGGLGDPGSLFLEGODQOHDPDTSDSLJEDDQSGSYASTSS 2740  
 DB 2163 PFLAREBSALNPGQPPGGLGDPGSLFLEGODQOHDPDTSDSLJEDDQSGSYASTSS 2222  
 QY 2741 DSEEBEEREBBAAPGEGWDSLIGPGBERLPLHSTPKDGGPGGKAPWPGDFGTAK 2800  
 DB 2223 DSEEBEEREBBAAPGEGWDSLIGPGBERLPLHSTPKDGGPGGKAPWPGDFGTAK 2282  
 QY 2801 SSGNGAPREERLRENDALSRGSLGPLPGSSAOPHKGLIKKCLPTISEKSLRLPLEQ 2860  
 DB 2283 SSGNGAPREERLRENDALSRGSLGPLPGSSAOPHKGLIKKCLPTISEKSLRLPLEQ 2342  
 QY 2861 CTGSSRGSASAGSGRGPPTPPRQSLQOBOLNGWPIAMSIKAGTVDESSGSEFLPFN 2920  
 DB 2343 CTGSSRGSASAGSGRGPPTPPRQSLQOBOLNGWPIAMSIKAGTVDESSGSEFLPFN 2402  
 QY 2921 FLH 2923  
 DB 2403 FLH 2405

XX DE Human calcitonin receptor.  
 XX XX  
 KW Calcitonin receptor; HCRPR6; G-protein coupled receptor; human; infection; migraine; cancer; anorexia; bulimia; asthma; allergy; Parkinson's disease; acute heart failure; hypertension; hypertension; urinary retention; osteoporosis; angina pectoris; myocardial infarction; ulcer; benign prostatic hypertrophy; psychosis; neurological diseases; anxiety; schizophrenia; manic depression; delirium; dementia; mental retardation; dyskinesia; Huntington's disease; Gilles de la Tourette syndrome; diagnosis; therapy.  
 KW KW  
 OS Homo sapiens.  
 XX XX  
 XX XX  
 FH Key  
 FT Peptide  
 FT 1..16  
 FT /label= Sig\_peptide  
 FT 1..18  
 FT /label= Extracellular  
 FT /note= "Claim 10"  
 FT 20..49  
 FT /label= Transmembrane\_domain-1  
 FT /note= "Claim 10"  
 FT 50..59  
 FT /label= Intracellular  
 FT /note= "Claim 10"  
 FT 60..80  
 FT /label= Transmembrane\_domain-2  
 FT /note= "Claim 10"  
 FT 82..109  
 FT /label= Transmembrane\_domain-3  
 FT /note= "Claim 10"  
 FT 110..125  
 FT /label= Intracellular  
 FT /note= "Claim 10"  
 FT 126..147  
 FT /label= Transmembrane\_domain-4  
 FT /note= "Claim 10"  
 FT 148..159  
 FT /label= Extracellular  
 FT /note= "Claim 10"  
 FT 160..187  
 FT /label= Transmembrane\_domain-5  
 FT /note= "Claim 10"  
 FT 188..209  
 FT /label= Intracellular  
 FT /note= "Claim 10"  
 FT 209..230  
 FT /label= Transmembrane\_domain-6  
 FT /note= "Claim 10"  
 FT 231..236  
 FT /label= Extracellular  
 FT /note= "Claim 10"  
 FT 237..258  
 FT /label= Transmembrane\_domain-7  
 FT /note= "Claim 10"  
 FT 259..560  
 FT /label= Intracellular  
 FT /note= "Claim 10"  
 FT 1..18  
 FT /label= Extracellular  
 FT /note= "Claim 10"  
 FT 209821242-A1.  
 PN WO9821242-A1.  
 XX 22-MAY-1998.  
 XX 14-NOV-1997; 97WO-US21330.  
 XX 15-NOV-1996; 96US-0030934.  
 XX PR  
 XX PA (HUMA-) HUMAN GENOME SCI INC.

XX LI Y, Ruben SM, Soppet DR;  
 PI MPI; 1998-297869/26.  
 DR N-PSDB; AAV07219.  
 XX  
 PT New nucleic acid encoding calcitonin receptor or its fragments -  
 useful for diagnosis and treatment of infections, cancer, allergy or  
 neurological disease  
 PT  
 PS Claim 1; Fig 1A-B; 88pp; English.  
 XX  
 CC This polypeptide comprises a novel human calcitonin receptor (CR),  
 CC which is a member of the G-protein coupled receptor superfamily.  
 CC Its amino acid sequence was deduced from a human cerebellum cDNA  
 CC clone (see AAV07219). The receptor is about 21% identical to rat  
 CC CR and about 31% identical to hormone receptor EMRI. CR  
 CC polypeptides, including extracellular, intracellular and  
 CC transmembrane regions, as well as epitope-bearing portions, are  
 CC claimed, as are isolated nucleic acids, recombinant vectors, host  
 CC cells and antibodies. Conditions related to under-expression or  
 CC over-expression of calcitonin receptor are treated by administering  
 CC an (ant)agonist of the receptor. Specified conditions are infections  
 CC (bacterial, fungal, protozoal or viral, particularly HIV-1 or  
 CC HIV-2), pain, e.g. migraine, cancer, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypertension, hypotension,  
 CC urinary retention, osteoporosis, angina pectoris, myocardial  
 CC infarction, ulcer, allergy, benign prostatic hypertrophy, psychotic  
 CC and neurological diseases (anxiety, schizophrenia), manic depression,  
 CC delirium, dementia and severe mental retardation) and dyskinesia  
 CC (e.g. Huntington's disease or Gilles de la Tourette syndrome) (all  
 CC claimed). Host cells are used to express recombinant calcitonin  
 CC receptor polypeptides which are used as mol.wt. markers, as  
 CC immunogens, for isolation of interacting proteins (or their genes)  
 CC and to screen for therapeutic (ant)agonists.  
 CC  
 XX  
 SQ Sequence 568 AA;  
 Query Match 18.3%; Score 534; DB 19; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2388 LAALLTFPLTLRLIRSNQHGIRMLTALGLAOLVLLGNODLPACVIALILH 2447  
 DB 33 LAALLTFPLTLRLIRSNQHGIRMLTALGLAOLVLLGNODLPACVIALILH 92  
 QY 2448 FLYLCTPSMALLLEALHLYRALTEVRDVNTGPMRFYVLGNGVPAFITGLAVGLDPBGYGN 2507  
 DB 93 FLYLCTPSMALLLEALHLYRALTEVRDVNTGPMRFYVLGNGVPAFITGLAVGLDPBGYGN 152  
 QY 2508 PDRGWLSTIYDTLWISFAGPVAFVAVSMVFLYIIAARASCAARQGEPEKGPVSGLOPSFA 2567  
 DB 153 PDRGWLSTIYDTLWISFAGPVAFVAVSMVFLYIIAARASCAARQGEPEKGPVSGLOPSFA 212  
 QY 2568 VLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSYVVLSEKVKALKACSRK 2627  
 DB 213 VLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSYVVLSEKVKALKACSRK 272  
 QY 2628 PSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGSLHSTSRSGSQPSYIFPLLRER 2687  
 DB 273 PSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGSLHSTSRSGSQPSYIFPLLRER 332  
 QY 2688 SAANPGCGPGLDPSGLFLBEGDQOQHDPTDSDSDLSLEDDSGSYASTHSSDSEERER 2747  
 DB 333 SAANPGCGPGLDPSGLFLBEGDQOQHDPTDSDSDLSLEDDSGSYASTHSSDSEERER 392  
 QY 2748 EEEBEAFAPEGQMDSLGPGAEKRLPLHSTPKDGGPGKAPWPGDGTAKSSSGNGAP 2807  
 DB 393 EEEBEAFAPEGQMDSLGPGAEKRLPLHSTPKDGGPGKAPWPGDGTAKSSSGNGAP 452  
 QY 2808 EERLRNGDALSRGSLGAPLPGSSAQPHKGLKKKCLPTISKSSLLRLPLEQCTGSSRG 2867  
 DB 453 EERLRNGDALSRGSLGAPLPGSSAQPHKGLKKKCLPTISKSSLLRLPLEQCTGSSRG 512

QY 2868 SSASBSRGPPPPRPPRQSLQELNGVPMIAMSIAKAGTVDESSGSBELPFNF 2921  
 DB 513 SSASBSRGPPPPRPPRQSLQELNGVPMIAMSIAKAGTVDESSGSBELPFNF 566  
 RESULT 10  
 ABJ37074  
 ID ABJ37074 standard; Protein; 565 AA.  
 XX  
 AC ABJ37074;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Human breast cancer / ovarian cancer related protein #50.  
 XX  
 KW Human; cytostatic; breast cancer; ovarian cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003000012-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 21-JUN-2002; 2002WO-US19773.  
 XX  
 PR 21-JUN-2001; 2001US-300159P.  
 XX  
 PR 27-JUN-2001; 2001US-301351P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Velby OP;  
 XX  
 DR MPI; 2003-267848/26.  
 DR N-PSDB; ABT31943.  
 XX  
 PT Determining the presence of breast cancer in an individual, involves  
 PT using specific polynucleotide markers -  
 XX  
 PS Disclosure; Page 219-220; 233pp; English.  
 XX  
 CC The invention comprises a method for assessing whether a patient is  
 CC afflicted with breast cancer or ovarian cancer. The method involves the  
 CC use of specific DNA markers. The method of the invention is useful in the  
 CC detection and treatment of ovarian and breast cancer. Amino acid  
 CC sequences ABJ37025 - ABJ37080 represent human breast/ovarian cancer-  
 CC related proteins.  
 CC  
 XX  
 SQ Sequence 565 AA;  
 Query Match 16.6%; Score 485; DB 24; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2430 INQADLPFACVIALILHLYLCTPSMALLLEALHLYRALTEVRDVNTGPMRFYVLGNGV 2489  
 DB 74 INQADLPFACVIALILHLYLCTPSMALLLEALHLYRALTEVRDVNTGPMRFYVLGNGV 133  
 QY 2490 PAFITGLAVGLDPBGYGNPFCMLSTIYDTLWISFAGPVAFVAVSMVFLYIIAARASCAAO 2549  
 DB 134 PAFITGLAVGLDPBGYGNPFCMLSTIYDTLWISFAGPVAFVAVSMVFLYIIAARASCAAO 193  
 QY 2550 RQGEPEKGPVSGLOPSFAVLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSY 2609  
 DB 194 RQGEPEKGPVSGLOPSFAVLLLSATWTLALLSVNSDTLLFHYLFATCNCTGPGPIFLSY 253  
 QY 2610 VVLSEKVKALKACSRKSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGSLHST 2669  
 DB 254 VVLSEKVKALKACSRKSPDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGSLHST 313  
 QY 2670 SRSQKQPSYIFPLLRERBALNPGGPPGLDPSGLFLBEGDQOQHDPTDSDSDLSLEDD 2729  
 DB 314 SRSQKQPSYIFPLLRERBALNPGGPPGLDPSGLFLBEGDQOQHDPTDSDSDLSLEDD 373



QY 2730 GGGSYASTHSDDSEEEBAAAPGQWDSILGPAERLPLHSTPKDGGPGKAP 2789  
 DB 374 QGGSYASTHSDDSEEEBAAAPGQWDSILGPAERLPLHSTPKDGGPGKAP 433  
 QY 2790 WPGDFTTAKSSGNGAPBEERLRENDALSRGSLGPLPGSSAQPHKGLKCKLPTISE 2849  
 DB 434 WPGDFTTAKSSGNGAPBEERLRENDALSRGSLGPLPGSSAQPHKGLKCKLPTISE 493  
 QY 2850 KSLRLPLPQCTGSSRGSSASBGRGPPPPRPPQSLQOLNGVMPITAMSTKATVDE 2909  
 DB 499 KSLRLPLPQCTGSSRGSSASBGRGPPPPRPPQSLQOLNGVMPITAMSTKATVDE 553  
 QY 2910 DSSGS 2914  
 DB 554 DSSGS 558

RESULT 11  
 ABB89665  
 ID ABB89665 standard; Protein; 568 AA.  
 AC ABB89665;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 2041.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 OS Homo sapiens.  
 XX  
 FN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 DR WPI; 2002-122018/16.  
 DR N-PSDB; ABL90074.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 11; SEQ ID NO 2041; 2081bp + Sequence Listing; English.

The invention relates to novel genes (AB89449-AB10853) and proteins (AB89040-AB89044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SQ Sequence 568 AA;  
 Query Match 9.6%; Score 281; DB 23; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-249;  
 Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2556 TCNCTGPPFIFLSTVYLSKVRKALKACSRKPSDPALTTKSTTSSTNCSPYADGR 2655  
 DB 241 TCNCTGPPFIFLSTVYLSKVRKALKACSRKPSDPALTTKSTTSSTNCSPYADGR 300  
 QY 2656 YQPGDSAGLSHSTRSGKSPYIFLLRBSALNPGQPGGLDPSLPLEGDDQDHD 2715  
 DB 301 YQPGDSAGLSHSTRSGKSPYIFLLRBSALNPGQPGGLDPSLPLEGDDQDHD 360  
 QY 2716 PDTSDSDLSLEDDQSGSYASTHSDDSEEEBAAAPGQWDSILGPAERLPLH 2775  
 DB 361 PDTSDSDLSLEDDQSGSYASTHSDDSEEEBAAAPGQWDSILGPAERLPLH 420  
 QY 2776 STPKDGGPGKAPWPGDFTTAKSSGNGAPBEERLRENDALSRGSLGPLPGSSAQPH 2835  
 DB 421 STPKDGGPGKAPWPGDFTTAKSSGNGAPBEERLRENDALSRGSLGPLPGSSAQPH 480  
 QY 2836 KGLKCKLPTISREKSLRLPLRQCTGSSRGSSASBGRG 2876  
 DB 481 KGLKCKLPTISREKSLRLPLRQCTGSSRGSSASBGRG 521

RESULT 12  
 AAB56721  
 ID AAB56721 standard; Protein; 717 AA.  
 AC AAB56721;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Human prostate cancer antigen protein sequence SEQ ID NO.1299.  
 XX  
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
 KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;  
 KW vulnery; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease.  
 OS Homo sapiens.  
 XX  
 FN WO200055174-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05988.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI (ROSE/) ROSEN C A.  
 DR Rosen CA, Ruben SM;  
 DR WPI; 2000-587513/55.  
 DR N-PSDB; AAF15924.  
 XX  
 PT Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -  
 XX  
 PS Claim 11; Page 1720-1723; 2338bp; English.

AAF15566 to AAF16505 encode the human prostate cancer associated



XX DE Human expressed polypeptide SEQ ID NO 43.  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;  
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX KW neurological disease; infection; human; secreted protein.  
XX OS Homo sapiens.  
XX PN WO200155167-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01319.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 26-SEP-2000; 2000US-0235484.  
XX PR 01-DEC-2000; 2000US-0250160.  
XX PR 05-DEC-2000; 2000US-0251988.  
XX PR 06-DEC-2000; 2000US-0251479.  
XX PR 08-DEC-2000; 2000US-0251990.  
XX PR 11-DEC-2000; 2000US-0254097.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WI WIPI; 2001-465559/50.  
XX DR N-PSDB; AA199622.  
XX PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
XX PT diagnose diseases or disorders associated with polypeptides and for  
XX PT treating autoimmune diseases e.g., multiple sclerosis, rheumatoid  
XX PT arthritis -  
XX PS Claim 11; SEQ ID NO 43; 504bp + Sequence Listing; English.  
XX CC The invention relates to novel genes (AA199614-AA199654) and proteins  
XX CC (AB03062-AB03068) useful for preventing, treating or ameliorating  
XX CC medical conditions e.g. by protein or gene therapy. The genes are  
XX CC isolated from a range of human tissues disclosed in the specification.  
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX CC and parasitic infections.  
XX CC Note: The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 219 AA;  
Query Match 4.0%; Score 118; DB 22; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.2e-99;  
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
567 ELDREVDVPSRGVARDGCTALTRASASVTVLDVNNNNPTFPQPEVTVLNDPAAG 626  
Db 1 ELDREVDVPSRGVARDGCTALTRASASVTVLDVNNNNPTFPQPEVTVLNDPAAG 60  
Qy 627 TSVVTVSAVDRAHSAVITQTISGNTNRFSTISGSGGLVSLALPLDKLEROVYLA 684  
Db 61 TSVVTVSAVDRAHSAVITQTISGNTNRFSTISGSGGLVSLALPLDKLEROVYLA 118

RESULT 15  
AAU19681  
ID AAU19681 standard; Protein; 219 AA.  
XX AC AAU19681;  
XX DT 06-DEC-2001 (first entry)  
XX DE Human novel extracellular matrix protein, Seq ID No 331.  
XX KW Human; secreted extracellular matrix protein; immunomodulatory;  
XX KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;  
XX KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytosatic;  
XX KW antialzheimer; immune/autoimmune disease; HIV infection; anaemia;  
XX KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
XX KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
XX KW Senary syndrome; Gaucher's disease; neurological diseases;  
XX KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
XX KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
XX KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX OS Homo sapiens.  
XX PN WO200155368-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01348.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 11-JUL-2000; 2000US-0216880.  
XX PR 14-JUL-2000; 2000US-0217487.  
XX PR 14-JUL-2000; 2000US-0217496.  
XX PR 26-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 14-AUG-2000; 2000US-0220964.  
XX PR 14-AUG-2000; 2000US-0224518.  
XX PR 14-AUG-2000; 2000US-0224519.  
XX PR 14-AUG-2000; 2000US-0225213.  
XX PR 14-AUG-2000; 2000US-0225214.  
XX PR 14-AUG-2000; 2000US-0225266.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225268.  
XX PR 14-AUG-2000; 2000US-0225270.  
XX PR 14-AUG-2000; 2000US-0225447.  
XX PR 14-AUG-2000; 2000US-0225757.  
XX PR 14-AUG-2000; 2000US-0225758.  
XX PR 14-AUG-2000; 2000US-0225759.  
XX PR 18-AUG-2000; 2000US-0226279.  
XX PR 22-AUG-2000; 2000US-0226681.  
XX PR 22-AUG-2000; 2000US-0226868.  
XX PR 23-AUG-2000; 2000US-0227182.  
XX PR 23-AUG-2000; 2000US-0227009.  
XX PR 30-AUG-2000; 2000US-0228924.  
XX PR 01-SEP-2000; 2000US-0229287.  
XX PR 01-SEP-2000; 2000US-0229343.  
XX PR 01-SEP-2000; 2000US-0229344.  
XX PR 01-SEP-2000; 2000US-0229345.  
XX PR 05-SEP-2000; 2000US-0229509.  
XX PR 05-SEP-2000; 2000US-0229513.  
XX PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234397.  
 PR 25-SEP-2000; 2000US-0234398.  
 PR 26-SEP-2000; 2000US-0234584.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251859.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.

## (HUMA-) HUMAN GENOME SCI INC.

XX  
 PI Rosen CA, Barash SC, Ruben SM,  
 XX WPI; 2001-465572/50.  
 DR N-PSDB; AAG31252.  
 XX

PT Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 PS  
 XX Claim 11; SEQ ID No 331; 577pp; English.

CC The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
 CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation.

Query Match 4.0%; Score 118; DB 22; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.2e-99; Indels 0; Gaps 0;

Matches 118; Conservative 0; Mismatches 0;

QY 567 ELDREKVDYFSGVGRARDHGTPTALTSASVSVTVDVNDNNPTFPQPEYTVRLNEDAAVG 626  
 DB 1 ELDREKVDYFSGVGRARDHGTPTALTSASVSVTVDVNDNNPTFPQPEYTVRLNEDAAVG 60  
 QY 627 TSVTVTSADVDRASHVITTYQITSGNTRNRFSTSGGGGLYSALPLPYKLERQYVLA 684  
 DB 61 TSVTVTSADVDRASHVITTYQITSGNTRNRFSTSGGGGLYSALPLPYKLERQYVLA 118

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 11, 2004, 16:08:16 ; Search time 67 Seconds

(without alignments)  
9134.693 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923  
Sequence: 1 MESPATGVLPPTPPPLLL.....AGTVDEDSGSEPLFPNPLH 2923

Scoring table: OLIGO

Gapop 60.0 , Gapept 60.0

Searched: 801455 seqs, 20938283 residues

Word size : 0

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

```
1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubppa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	2923	11	US-09-788-711A-4
2	2923	100.0	2923	11	US-09-916-849A-3
3	2923	100.0	2923	12	US-10-174-677-29
4	2923	100.0	2923	12	US-10-120-801-53
5	2923	100.0	2923	12	US-10-292-798-932
6	2923	100.0	2923	15	US-10-225-567A-524
7	2837	97.1	2956	9	US-09-788-711A-2
8	534	18.3	568	9	US-09-843-856-2
9	485	16.6	565	15	US-10-176-847-100
10	281	9.6	568	12	US-10-264-237-2041
11	281	9.6	717	10	US-09-925-300-1299
12	179	6.1	646	12	US-10-017-161-1096
13	118	4.0	219	9	US-09-764-870-331
14	118	4.0	219	15	US-10-125-540-331
15	100	3.4	111	9	US-09-764-870-479

16	100	3.4	111	15	US-10-125-540-479	Sequence 479, App
17	82	2.8	96	9	US-09-764-853-659	Sequence 659, App
18	82	2.8	96	9	US-09-764-898-286	Sequence 286, App
19	82	2.8	96	11	US-09-764-881-94	Sequence 94, App1
20	82	2.8	96	12	US-10-242-747-94	Sequence 94, App1
21	82	2.8	96	15	US-10-073-865-108	Sequence 108, App
22	59	2.0	174	12	US-10-017-161-1688	Sequence 1688, App
23	42	1.4	141	9	US-09-764-870-335	Sequence 335, App
24	42	1.4	141	15	US-10-125-540-335	Sequence 335, App
25	23	0.8	3034	9	US-09-737-149-25	Sequence 25, App1
26	23	0.8	3034	9	US-09-737-149-30	Sequence 30, App1
27	23	0.8	3034	12	US-10-131-409-70	Sequence 70, App1
28	23	0.8	3034	12	US-10-120-801-52	Sequence 52, App1
29	23	0.8	3034	12	US-10-150-811-70	Sequence 70, App1
30	21	0.7	44	12	US-09-880-573-16	Sequence 16, App1
31	20	0.7	20	11	US-09-916-849A-11	Sequence 11, App1
32	20	0.7	20	15	US-10-225-567A-1865	Sequence 1865, App
33	20	0.7	20	15	US-10-225-567A-1887	Sequence 1887, App
34	20	0.7	1713	9	US-09-737-149-27	Sequence 27, App1
35	20	0.7	3014	9	US-09-737-149-2	Sequence 2, App1
36	20	0.7	3014	12	US-10-241-220-107	Sequence 107, App
37	20	0.7	3014	12	US-10-174-677-77	Sequence 77, App1
38	20	0.7	3014	12	US-10-295-027-750	Sequence 750, App
39	20	0.7	3014	12	US-10-240-145-63	Sequence 63, App1
40	20	0.7	3014	15	US-10-225-567A-444	Sequence 444, App
41	20	0.7	3028	12	US-10-131-409-14	Sequence 14, App1
42	20	0.7	3028	12	US-10-131-409-69	Sequence 69, App1
43	20	0.7	3028	12	US-10-150-811-14	Sequence 14, App1
44	20	0.7	3028	12	US-10-150-811-69	Sequence 69, App1
45	19	0.7	19	15	US-10-225-567A-1865	Sequence 1865, App

## ALIGNMENTS

```
RESULT 1
US-09-788-711A-4
; Sequence 4, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GR-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 2923
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-788-711A-4
```

Query Match 100.0%; Score 2923; DB 9; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MESPATGVLPPTPPPLLLLLLPPLLDQVPCRSLSGRGSSGACAPMGLCP	60
DB	1	MESPATGVLPPTPPPLLLLLLPPLLDQVPCRSLSGRGSSGACAPMGLCP	60
QY	61	SASNTMLTSCRDAETLVTGLVPHDGLRWVCESBAHITPLPAPBCGCPMSCLTIG	120
DB	61	SASNTMLTSCRDAETLVTGLVPHDGLRWVCESBAHITPLPAPBCGCPMSCLTIG	120
QY	121	GHLSFGKLTLPREHPCIKAPLRCSCTLAQAPRLRAGERSPEBSLGGRRRNNTAPQ	180
DB	121	GHLSFGKLTLPREHPCIKAPLRCSCTLAQAPRLRAGERSPEBSLGGRRRNNTAPQ	180
QY	181	FGPPSYQATVPENQAGTVPASIRAIIDDEGAGLETMTALPFSRSNQPSLDPVIGA	240

Dh 181 FQPSYQATVRENOPAGTVPVSLRAIDDEGEAGRLLEYMDALPDRSNOFSLDPVTGA 240  
Qy VTTAEELDRKSTHFRVTAQDHGMPRRSALATITLVTDTNDHPVFEQOEKESLRE 300  
Dh 241 VTTAEELDRKSTHFRVTAQDHGMPRRSALATITLVTDTNDHPVFEQOEKESLRE 300  
Qy 301 NLEVGVEVLVTRATDGAAPPNANILYRLLESGSGSPSEVFEIDPRSGVIRTRGVDBEEV 360  
Dh 301 NLEVGVEVLVTRATDGAAPPNANILYRLLESGSGSPSEVFEIDPRSGVIRTRGVDBEEV 360  
Qy 361 ESYQVLTBASDQDGPGRSTTAAPVLSVEDDNNAPQFSKRVVQVREBVTGAYLR 420  
Dh 361 ESYQVLTBASDQDGPGRSTTAAPVLSVEDDNNAPQFSKRVVQVREBVTGAYLR 420  
Qy 421 VTTASDRKSGSNAAVHYHISMSGNAGQFYLDAQTALDVSPFLDYETTKETYLTRAADGG 480  
Dh 421 VTTASDRKSGSNAAVHYHISMSGNAGQFYLDAQTALDVSPFLDYETTKETYLTRAADGG 480  
Qy 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFGATVLESVPLGYLVLAHQADADAGNARL 540  
Dh 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFGATVLESVPLGYLVLAHQADADAGNARL 540  
Qy 541 EYRLAGVGHDPFTINNGTGMISVAELDRREVDYFSGVEARDHGTPALTAASVSTV 600  
Dh 541 EYRLAGVGHDPFTINNGTGMISVAELDRREVDYFSGVEARDHGTPALTAASVSTV 600  
Qy 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVVTSVAVDASHSVITYQITSGNTRRFSITS 660  
Dh 601 LDVNDNNPTFTQPEYTVRLNEDAAVGTSVVTSVAVDASHSVITYQITSGNTRRFSITS 660  
Qy 661 QSGGGLVSLALPLDYKLEROYVLAVTASDGTRODTAQIVNVNTDANTHRPVQSHYTVN 720  
Dh 661 QSGGGLVSLALPLDYKLEROYVLAVTASDGTRODTAQIVNVNTDANTHRPVQSHYTVN 720  
Qy 721 VNEBRPAGTTVVLSATDEDEGENARITYEMEDSIPOFRIDADGATTOAELDYEBQVS 780  
Dh 721 VNEBRPAGTTVVLSATDEDEGENARITYEMEDSIPOFRIDADGATTOAELDYEBQVS 780  
Qy 781 YTLAITARNDNGIPQKSDPTTYLEILVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLQISAT 840  
Dh 781 YTLAITARNDNGIPQKSDPTTYLEILVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLQISAT 840  
Qy 841 DROSGLNGRYFTYTOGGDDGDGDFIVESTSGIVTTLRLDRBNVAQVYLRAVADKMP 900  
Dh 841 DROSGLNGRYFTYTOGGDDGDGDFIVESTSGIVTTLRLDRBNVAQVYLRAVADKMP 900  
Qy 901 ARTPMEVTVVLYDNDNPPVPEODEPVPVEBNSPIGLAVARYATPDDEGTNAQIMYOI 960  
Dh 901 ARTPMEVTVVLYDNDNPPVPEODEPVPVEBNSPIGLAVARYATPDDEGTNAQIMYOI 960  
Qy 961 VEGNIPREVFOIDIFSGBELTALVDLYEDREPEYVLVIQATSAPLVSRAVTVHRLDRNDP 1020  
Dh 961 VEGNIPREVFOIDIFSGBELTALVDLYEDREPEYVLVIQATSAPLVSRAVTVHRLDRNDP 1020  
Qy 1021 PVLGNFELLRNNTYTNRSSSPFGGAIQGVPAHDPIDISLTSYBERGNEISLVILNASTG 1080  
Dh 1021 PVLGNFELLRNNTYTNRSSSPFGGAIQGVPAHDPIDISLTSYBERGNEISLVILNASTG 1080  
Qy 1081 ELKLSRLDNNRPLEAIVSVLVSDGVASVTAQCALRTIITDEMLTHSITLLEDMSPER 1140  
Dh 1081 ELKLSRLDNNRPLEAIVSVLVSDGVASVTAQCALRTIITDEMLTHSITLLEDMSPER 1140  
Qy 1141 FLSPBLGLFTQAAVATLATPPDHVVENVQRTDAPGSHILNVLSTVGQPGGCGPFL 1200  
Dh 1141 FLSPBLGLFTQAAVATLATPPDHVVENVQRTDAPGSHILNVLSTVGQPGGCGPFL 1200  
Qy 1201 PSEBLOERLYKINSILTAISAQRYLPPDDNICLAEPEBNYRMCVSVLRPSSAPFLTSSS 1260  
Dh 1201 PSEBLOERLYKINSILTAISAQRYLPPDDNICLAEPEBNYRMCVSVLRPSSAPFLTSSS 1260  
Qy 1261 VLFRRHPVGGRLCRCPGFTGVDCETEVLDLCTSRPCGPHGRCSRREGVTCCLCRDGYTG 1320  
Dh 1261 VLFRRHPVGGRLCRCPGFTGVDCETEVLDLCTSRPCGPHGRCSRREGVTCCLCRDGYTG 1320

Qy 1321 EHCEVSARSGRCTPGVCKNGTCVNLVGGFKDCPSGDPFEKPYCOVTTSPFAHSFTTF 1380  
Dh 1321 EHCEVSARSGRCTPGVCKNGTCVNLVGGFKDCPSGDPFEKPYCOVTTSPFAHSFTTF 1380  
Qy 1381 RGLRQRFHFTLALSPATKERDGLLLYNGRFNEKHDVFALEVIOBOVLTFSAGESSTTVS 1440  
Dh 1381 RGLRQRFHFTLALSPATKERDGLLLYNGRFNEKHDVFALEVIOBOVLTFSAGESSTTVS 1440  
Qy 1441 PVPVGVSDGQMTVQOLKYNKPLDGTGLPOGSSQKAAVTVVDGCDGVALARFQSVLG 1500  
Dh 1441 PVPVGVSDGQMTVQOLKYNKPLDGTGLPOGSSQKAAVTVVDGCDGVALARFQSVLG 1500  
Qy 1501 NYSCAAQGTGSKSKSLDLTGPLLGGVDPDLPSPFVVRMRQFVGCNRNLOVDSRHIDMAD 1560  
Dh 1501 NYSCAAQGTGSKSKSLDLTGPLLGGVDPDLPSPFVVRMRQFVGCNRNLOVDSRHIDMAD 1560  
Qy 1561 FIANNGTVPCCPKAKVCDNSNTCHNGGTCVQMDAPSCBCLPFGGKSCAQEMANPOHFL 1620  
Dh 1561 FIANNGTVPCCPKAKVCDNSNTCHNGGTCVQMDAPSCBCLPFGGKSCAQEMANPOHFL 1620  
Qy 1621 GSSLVAMHGLSLPISQPWYLSIMFRTRQADGVLLQAITGRSTTYQLBEGHVLSEGT 1680  
Dh 1621 GSSLVAMHGLSLPISQPWYLSIMFRTRQADGVLLQAITGRSTTYQLBEGHVLSEGT 1680  
Qy 1681 GLQASSLRLEPGRANDGWHHAQALAGASGPGHALLSFYDQOQRAEGNLGPRHGLHLS 1740  
Dh 1681 GLQASSLRLEPGRANDGWHHAQALAGASGPGHALLSFYDQOQRAEGNLGPRHGLHLS 1740  
Qy 1741 NITVGGI PGAGVANGFRGCIQGVVSDTPBSGNSLDPHSGESINVEQCSLPDPCDSN 1800  
Dh 1741 NITVGGI PGAGVANGFRGCIQGVVSDTPBSGNSLDPHSGESINVEQCSLPDPCDSN 1800  
Qy 1801 PCPANSYCSNDMDYSQSCDPPGYGDNCTNVCIDLANCEHOSVCTRPSAPHGTGCPBN 1860  
Dh 1801 PCPANSYCSNDMDYSQSCDPPGYGDNCTNVCIDLANCEHOSVCTRPSAPHGTGCPBN 1860  
Qy 1861 YLGPYCETRIDQCPRGMMGHPTGCPNCNCDVSKGPPDNCNKTSGECHKENHRRPGSPT 1920  
Dh 1861 YLGPYCETRIDQCPRGMMGHPTGCPNCNCDVSKGPPDNCNKTSGECHKENHRRPGSPT 1920  
Qy 1921 CLLCDCTPPGSLSRVCDPEBQCPCKRGVYIGRCQRCNDNPAEVTNNGCBVATDSCPRAI 1980  
Dh 1921 CLLCDCTPPGSLSRVCDPEBQCPCKRGVYIGRCQRCNDNPAEVTNNGCBVATDSCPRAI 1980  
Qy 1981 EAGIMPRTRFGLPAAAPCPKSGFGTAVRHCEHRGMLPNNLFNCTSIIFSELKGFABRL 2040  
Dh 1981 EAGIMPRTRFGLPAAAPCPKSGFGTAVRHCEHRGMLPNNLFNCTSIIFSELKGFABRL 2040  
Qy 2041 QRNESGLDSGRSQOLALLRNATQHTAGYFGSDVKAVALATRLAHBSTORFGLSATQ 2100  
Dh 2041 QRNESGLDSGRSQOLALLRNATQHTAGYFGSDVKAVALATRLAHBSTORFGLSATQ 2100  
Qy 2101 DVHFTENLLRVGSALLDTANKRHEMLIOQTBEGTALLQHTERYAALAAQNNRHLYLSFP 2160  
Dh 2101 DVHFTENLLRVGSALLDTANKRHEMLIOQTBEGTALLQHTERYAALAAQNNRHLYLSFP 2160  
Qy 2161 TIVTPNIVSVVRLDKGNFAGAKLPRYBALRSGOPDLETTYVLPESVRETPPVVRPAG 2220  
Dh 2161 TIVTPNIVSVVRLDKGNFAGAKLPRYBALRSGOPDLETTYVLPESVRETPPVVRPAG 2220  
Qy 2221 PGBAQBEEELARRORRHEPLSGRAVASVYIYRTLAGLLPHNYDPDKSLRVPKPIINT 2280  
Dh 2221 PGBAQBEEELARRORRHEPLSGRAVASVYIYRTLAGLLPHNYDPDKSLRVPKPIINT 2280  
Qy 2281 PUVYSIVHDEBELPALDKPVTVOFRILETERKTPICVFNHNSILVSGGWSARGGE 2340  
Dh 2281 PUVYSIVHDEBELPALDKPVTVOFRILETERKTPICVFNHNSILVSGGWSARGGE 2340  
Qy 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLLAALLTFPFLTL 2400  
Dh 2341 VVFRNESHVSCQNMHTSPAVLMDVSRRENGEILPLKTLTYVALGVTLLAALLTFPFLTL 2400



QY 2401 LKILNSNGGIRRNLTALGLAQLVFLGINQADLPACTVIAIILHFLYLCFSSWALLE 2460  
 Db 2401 LKILNSNGGIRRNLTALGLAQLVFLGINQADLPACTVIAIILHFLYLCFSSWALLE 2460  
 QY 2461 ALHLVALLEVRDVTNGMRFYMLGMPAPFTTGLAVGLDPRGYNPDPCHLSITDTLI 2520  
 Db 2461 ALHLVALLEVRDVTNGMRFYMLGMPAPFTTGLAVGLDPRGYNPDPCHLSITDTLI 2520  
 QY 2521 WSPAGVAVAVSVSVFLYILAAASCAARQCFEKKGPVSGLOPSPAVILLISATWILAL 2580  
 Db 2521 WSPAGVAVAVSVSVFLYILAAASCAARQCFEKKGPVSGLOPSPAVILLISATWILAL 2580  
 QY 2581 LSVNSSTLLPHLIPATCNQIGPPFLSYVTSKEYRAKAKLACSRKPPDPALTTKSTL 2640  
 Db 2581 LSVNSSTLLPHLIPATCNQIGPPFLSYVTSKEYRAKAKLACSRKPPDPALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLOPYGDSAGSLHSTRSGKSPSYTPILLREBSALNPGQGPGLG 2700  
 Db 2641 TSSYNCPSPYADGRLOPYGDSAGSLHSTRSGKSPSYTPILLREBSALNPGQGPGLG 2700  
 QY 2701 DPGSLFLBQDQDHDPTDSDSLSDQSGSYASTHSDSEEESEEESEEAAPFGEG 2760  
 Db 2701 DPGSLFLBQDQDHDPTDSDSLSDQSGSYASTHSDSEEESEEESEEAAPFGEG 2760  
 QY 2761 WPSLCPGAEPLPLHSTPDGPGPKAPWPGDFTTAKSSGNGAPDEERLRENGDALSR 2820  
 Db 2761 WPSLCPGAEPLPLHSTPDGPGPKAPWPGDFTTAKSSGNGAPDEERLRENGDALSR 2820  
 QY 2821 EBSLGLPSSAOPHGLIKKCLPTISEKSLRLPLFOCTGSSSGSSASBSRGPP 2880  
 Db 2821 EBSLGLPSSAOPHGLIKKCLPTISEKSLRLPLFOCTGSSSGSSASBSRGPP 2880  
 QY 2881 RPPRQSLQEOULNGVPIAMSIKAGTVDEDSGSEFLFFNFIH 2923  
 Db 2881 RPPRQSLQEOULNGVPIAMSIKAGTVDEDSGSEFLFFNFIH 2923

RESULT 2  
 US-09-916-849A-3  
 ; Sequence 3, Application US/0916849A  
 ; Publication No. US20030086934A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bostein, et al.  
 ; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents  
 ; TITLE OF INVENTION: Uses Thereof  
 ; FILE REFERENCE: 2002850-0024  
 ; CURRENT APPLICATION NUMBER: US/09/916,849A  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 2923  
 ; TYPE: PRF  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Cadherin EGF  
 ; OTHER INFORMATION: IAG Seven Pass G-Type Receptor 2  
 US-09-916-849A-3

Query Match 100.0%; Score 2923; DB 11; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLIGDQVPCRSIGSRGSGSACA PMGMLCPS 60  
 Db 1 MSPPATGVLPPTPPPLLLLLLLLLLPPPLIGDQVPCRSIGSRGSGSACA PMGMLCPS 60  
 QY 61 SASNLMLYTSRCRDACTELTGHLVPHHDGIRVWCPESEAHIPLPAPAECCPMSCRLIGIG 120  
 Db 61 SASNLMLYTSRCRDACTELTGHLVPHHDGIRVWCPESEAHIPLPAPAECCPMSCRLIGIG 120  
 QY 121 GHLSPGKLTTPBEHCLKAPRLRCQSCCLAQAPGLRAGERSPEESLGRRRRKRVNTAQQ 180  
 Db 121 GHLSPGKLTTPBEHCLKAPRLRCQSCCLAQAPGLRAGERSPEESLGRRRRKRVNTAQQ 180

Db 121 GHLSPGKLTTPBEHCLKAPRLRCQSCCLAQAPGLRAGERSPEESLGRRRRKRVNTAQQ 180  
 QY 181 FQPPSYOATVPENOPAGTPVASTRAIDPDEGAGLEMTDALPDSRSNOFSLDPVTA 240  
 Db 181 FQPPSYOATVPENOPAGTPVASTRAIDPDEGAGLEMTDALPDSRSNOFSLDPVTA 240  
 QY 241 VTTAEELRETKSTVFRTAODHGMPPRSALATLTITLVTDNHD PVEQOEYKESLRE 300  
 Db 241 VTTAEELRETKSTVFRTAODHGMPPRSALATLTITLVTDNHD PVEQOEYKESLRE 300  
 QY 301 NIEVGYEVLTVATATGDAPPNANILYRLLEGSGSPSEVEFIDPDSGVRTGPPVDREY 360  
 Db 301 NIEVGYEVLTVATATGDAPPNANILYRLLEGSGSPSEVEFIDPDSGVRTGPPVDREY 360  
 QY 361 ESYOULTVEASDQGRDPGPSTTAAVFLSVEDNDNAPQSEKRYVQVQVEDVTPGAPVR 420  
 Db 361 ESYOULTVEASDQGRDPGPSTTAAVFLSVEDNDNAPQSEKRYVQVQVEDVTPGAPVR 420  
 QY 421 VTASDRDKGSNAVVAHYSIMSGNARQFYLDAGTALDVVSPLDYETTKXYTLRVAAQDGG 480  
 Db 421 VTASDRDKGSNAVVAHYSIMSGNARQFYLDAGTALDVVSPLDYETTKXYTLRVAAQDGG 480  
 QY 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGVLVHQAIDADAGDNAL 540  
 Db 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGVLVHQAIDADAGDNAL 540  
 QY 541 EYRLAGVGHDPFTLNNGTGMSVAEELDRREVDPYSGEVARDHGTPLTASASVTV 600  
 Db 541 EYRLAGVGHDPFTLNNGTGMSVAEELDRREVDPYSGEVARDHGTPLTASASVTV 600  
 QY 601 LDVNDNPPFTQPEXTVRLNEDAAVGTSVTVSAVDRDAHSVTVQITSGNTRNFSITS 660  
 Db 601 LDVNDNPPFTQPEXTVRLNEDAAVGTSVTVSAVDRDAHSVTVQITSGNTRNFSITS 660  
 QY 661 QSGGGLVSLALPLDYKLRQVYLAATASDQTRQDTRQVTVNTDANTHRPVPQSHYTN 720  
 Db 661 QSGGGLVSLALPLDYKLRQVYLAATASDQTRQDTRQVTVNTDANTHRPVPQSHYTN 720  
 QY 721 VNEBRPAGTIVVLSATDEDTGENARITYFMEDSIPQFRIDADGAVTTQAELEDQVS 780  
 Db 721 VNEBRPAGTIVVLSATDEDTGENARITYFMEDSIPQFRIDADGAVTTQAELEDQVS 780  
 QY 781 YTLAATARDNGIPQKSDTTYLEILVNDVNDNAPQFLRDSYQSGSYVEDVPPFTSVLQISAT 840  
 Db 781 YTLAATARDNGIPQKSDTTYLEILVNDVNDNAPQFLRDSYQSGSYVEDVPPFTSVLQISAT 840  
 QY 841 DRDSGLNGRVPTFGGSDGDDPFIVESTSGVRLRLDRDENNAQVYLAAYADKGMPP 900  
 Db 841 DRDSGLNGRVPTFGGSDGDDPFIVESTSGVRLRLDRDENNAQVYLAAYADKGMPP 900  
 QY 901 ARTPMETVTVLVDVNDNPPVFEQDEFDVFVEENSPIGLAVARVATDDEGNAQIMYQI 960  
 Db 901 ARTPMETVTVLVDVNDNPPVFEQDEFDVFVEENSPIGLAVARVATDDEGNAQIMYQI 960  
 QY 961 VEGNIPEYFOLDIFSGELTALVDLYEDRPREVYVLOATSAPLVSRATVHVALLDRNDP 1020  
 Db 961 VEGNIPEYFOLDIFSGELTALVDLYEDRPREVYVLOATSAPLVSRATVHVALLDRNDP 1020  
 QY 1021 PVLGNFELFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFERGNLSLYLNA5TG 1080  
 Db 1021 PVLGNFELFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFERGNLSLYLNA5TG 1080  
 QY 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTAQCALRVITIDEMLTHSITTLLEDMSPER 1140  
 Db 1081 ELKLSRALDNNRPLEAIVSVSDGVSVTAQCALRVITIDEMLTHSITTLLEDMSPER 1140  
 QY 1141 FLSPPLSLFLIOAVALTATPPDHVVVFVNOBTDAPGGHILNVSLSVQPPGPGGPPFL 1200  
 Db 1141 FLSPPLSLFLIOAVALTATPPDHVVVFVNOBTDAPGGHILNVSLSVQPPGPGGPPFL 1200  
 QY 1201 PSEBDQERLYLNRSILTAISAQRYLPDDNITCLAREPCENYKCVSVLRFDSAPPIASS 1260  
 Db 1201 PSEBDQERLYLNRSILTAISAQRYLPDDNITCLAREPCENYKCVSVLRFDSAPPIASS 1260

QY 1261 VLEPPIHPVGLRCRCPPGFTGPDYCEYVLDLCTSRPQCPHRCRSGRGVTLCDYTG 1320  
 DB 1261 VLEPPIHPVGLRCRCPPGFTGPDYCEYVLDLCTSRPQCPHRCRSGRGVTLCDYTG 1320  
 QY 1321 EHCEVSARSGRCTPGVCCKNGSTCVNLLVGGFKDCDPSGDEKPYCQVTTSPHASEITP 1380  
 DB 1321 EHCEVSARSGRCTPGVCCKNGSTCVNLLVGGFKDCDPSGDEKPYCQVTTSPHASEITP 1380  
 QY 1381 RGRORHPHFLTALSPATKEDGILLNNGRPNKHDPALEVOIOBOVOLTSSABESTTYS 1440  
 DB 1381 RGRORHPHFLTALSPATKEDGILLNNGRPNKHDPALEVOIOBOVOLTSSABESTTYS 1440  
 QY 1441 PFPVGVSDQGMHTVOLKYNKPLAQOTGLPGSPSEOKAVAVTVGDCDTVALRFGSVLG 1500  
 DB 1441 PFPVGVSDQGMHTVOLKYNKPLAQOTGLPGSPSEOKAVAVTVGDCDTVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQSSKKSIDLTPLLGSPDLPSFPVPRKQPVGCMRNLTQVDSRHIDMAD 1560  
 DB 1501 NYSCAAQGTQSSKKSIDLTPLLGSPDLPSFPVPRKQPVGCMRNLTQVDSRHIDMAD 1560  
 QY 1561 FIANNQVPGCPAKKQVCDSTNGCTCVNOMDASECEPLRGKSCAQMANNPQHL 1620  
 DB 1561 FIANNQVPGCPAKKQVCDSTNGCTCVNOMDASECEPLRGKSCAQMANNPQHL 1620  
 QY 1621 GSSLVAMHGLSLPIQPMWYLSMFRTRQADGVLLQAITRGRSTTLQLRGAHMLSYEGT 1680  
 DB 1621 GSSLVAMHGLSLPIQPMWYLSMFRTRQADGVLLQAITRGRSTTLQLRGAHMLSYEGT 1680  
 QY 1681 GLOASSILREPRANDMDHMAOLALGASGPGHAILSPDYGOORAGNIGPRLGLHLS 1740  
 DB 1681 GLOASSILREPRANDMDHMAOLALGASGPGHAILSPDYGOORAGNIGPRLGLHLS 1740  
 QY 1741 NITVGGIPGAGVANGFRGCLGVNRYSDTPEGVNSLDPSHGSINVEQCSLPDPDSN 1800  
 DB 1741 NITVGGIPGAGVANGFRGCLGVNRYSDTPEGVNSLDPSHGSINVEQCSLPDPDSN 1800  
 QY 1801 PCPANSYCSNDMDSYSCSDPGYGDNCYVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 DB 1801 PCPANSYCSNDMDSYSCSDPGYGDNCYVCDLNPCEHOSVCTRKSAPHGTYCECPN 1860  
 QY 1861 YLGPYCECTRIDOPCFRGMWGHPTGCPNCVSKGFPDCKNTSGECKENHRRPSPPT 1920  
 DB 1861 YLGPYCECTRIDOPCFRGMWGHPTGCPNCVSKGFPDCKNTSGECKENHRRPSPPT 1920  
 QY 1921 CLICDCYPTGSLRSVCPEDGQCPCKPGVIGROCDRCNDPFAVTTNGCEVNTDSCPRAT 1980  
 DB 1921 CLICDCYPTGSLRSVCPEDGQCPCKPGVIGROCDRCNDPFAVTTNGCEVNTDSCPRAT 1980  
 QY 1981 EAGIWPRTFGLPAAAPCEPKGSPGTAVRHCDEHGMPLNLFUNCTSITPSELKGFARL 2040  
 DB 1981 EAGIWPRTFGLPAAAPCEPKGSPGTAVRHCDEHGMPLNLFUNCTSITPSELKGFARL 2040  
 QY 2041 QRNESGIDSRSOQALLLNATQHTAGYFGSDYKAVAYOATRLLANESTORGGLSATQ 2100  
 DB 2041 QRNESGIDSRSOQALLLNATQHTAGYFGSDYKAVAYOATRLLANESTORGGLSATQ 2100  
 QY 2101 DVHFTETLAVGSAALLDTANKRHMLIQOTEGGTAMLLQHYEYASALAQMHTYISPF 2160  
 DB 2101 DVHFTETLAVGSAALLDTANKRHMLIQOTEGGTAMLLQHYEYASALAQMHTYISPF 2160  
 QY 2161 TIVTPNIVISVRLDKGNFAGAKLPRYEALRGEQPPDLETTVILPESVFRETPEVVPAG 2220  
 DB 2161 TIVTPNIVISVRLDKGNFAGAKLPRYEALRGEQPPDLETTVILPESVFRETPEVVPAG 2220  
 QY 2221 PGGAOBEELARORRHPBELSGEBAVASVITYRTLAGLPHNYDPDRSRARVPRPINT 2280  
 DB 2221 PGGAOBEELARORRHPBELSGEBAVASVITYRTLAGLPHNYDPDRSRARVPRPINT 2280  
 QY 2281 PVSISVHDBELLPRALDKPVTVQFRLLETERTKPICVFMNHSILVSGTGMSARGCE 2340  
 DB 2281 PVSISVHDBELLPRALDKPVTVQFRLLETERTKPICVFMNHSILVSGTGMSARGCE 2340

QY 2341 VVFNESHVSQCCNMTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFEFLT 2400  
 DB 2341 VVFNESHVSQCCNMTSPAVLMDVSRRENGEILPLKTLTYVALGVTLAALLTFEFLT 2400  
 QY 2401 LRLNSNOGIRRNLTALGLAQVFLGIGNADLPACTVIAILLHFLYLCFSPWALLE 2460  
 DB 2401 LRLNSNOGIRRNLTALGLAQVFLGIGNADLPACTVIAILLHFLYLCFSPWALLE 2460  
 QY 2461 ALHLVRALETVADVTGPMRFYVLMGMGPAPITGLAVGLDREGVGNPCWMLISTDTLI 2520  
 DB 2461 ALHLVRALETVADVTGPMRFYVLMGMGPAPITGLAVGLDREGVGNPCWMLISTDTLI 2520  
 QY 2521 WSPFAGVAVASVFLYTLAARASCAQROGFEKKGPVSGLOPFAVILLLSATVLLAL 2580  
 DB 2521 WSPFAGVAVASVFLYTLAARASCAQROGFEKKGPVSGLOPFAVILLLSATVLLAL 2580  
 QY 2581 LSVNSDTLLFHTLPATCNCIQPPIFLSYVLSKRYKALKACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLPATCNCIQPPIFLSYVLSKRYKALKACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKSQPSYIPFLRBSALNPGQGPGLG 2700  
 DB 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKSQPSYIPFLRBSALNPGQGPGLG 2700  
 QY 2701 DPGSLFLEGODQDHPDSDLSLKDQSGSYASTHSDSEEBEEREBEBAAPFGEQ 2760  
 DB 2701 DPGSLFLEGODQDHPDSDLSLKDQSGSYASTHSDSEEBEEREBEBAAPFGEQ 2760  
 QY 2761 WDSLAPGABRLPLHSTPDGPGPGKAPWPGDPTGTAKSSGNAPBERLBNGLALS 2820  
 DB 2761 WDSLAPGABRLPLHSTPDGPGPGKAPWPGDPTGTAKSSGNAPBERLBNGLALS 2820  
 QY 2821 EBSLGLPGSSAQPHKGLKXKCLPTISEKSLRLPLEQCGSSRSGSSRSGRGP 2880  
 DB 2821 EBSLGLPGSSAQPHKGLKXKCLPTISEKSLRLPLEQCGSSRSGSSRSGRGP 2880  
 QY 2881 RPPRQSLQEQLVGWPVIMSIAKGTVDSSGSEFLFENFLH 2923  
 DB 2881 RPPRQSLQEQLVGWPVIMSIAKGTVDSSGSEFLFENFLH 2923

RESULT 3  
 US-10-174-677-29  
 ; Sequence 29, Application US/10174677  
 ; Publication No. US20030190704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xie, Ting  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR  
 ; FILE REFERENCE: 40716(1P-012)  
 ; CURRENT APPLICATION NUMBER: US/10/174,677  
 ; CURRENT FILING DATE: 2002-06-19  
 ; NUMBER OF SEQ ID NOS: 117  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-174-677-29

Query Match 100.0%; Score 2923; DB 12; Length 2923;  
 Best Local Similarity 100.0%; Fred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSBPATGVLPPTPPPLLLLLLLLLLPPPLGQVGCPSRISGRSGSSGACAPMGWLCPS 60  
 DB 1 MSBPATGVLPPTPPPLLLLLLLLLLPPPLGQVGCPSRISGRSGSSGACAPMGWLCPS 60  
 QY 61 SASNMLYTSRCKDAGTBLTGHVPHHDLRYWCPSBAHILPLPAPBECGPMSCRLGIG 120  
 DB 61 SASNMLYTSRCKDAGTBLTGHVPHHDLRYWCPSBAHILPLPAPBECGPMSCRLGIG 120  
 QY 121 GHLSPGKTLTLEHPCCLAPRLRCOSCTLAAPGLRAGERSPEBSLGGRRKNVATAPO 180  
 DB 121 GHLSPGKTLTLEHPCCLAPRLRCOSCTLAAPGLRAGERSPEBSLGGRRKNVATAPO 180

Db 121 GHLISQKLTLEBHPCLAPRLKQSCKLAQAPGIRAGERSPEBSLGRRRRNVNTAQ 180  
 Qy 181 FQPSYQATVPENOPAGTVPASLARAIDPEBEGAGLEETMDLPDSRNQFSLDPVTGA 240  
 Db 181 FQPSYQATVPENOPAGTVPASLARAIDPEBEGAGLEETMDLPDSRNQFSLDPVTGA 240  
 Qy 241 VTTAEELDRETKSTVFRVTAODHGMRRSALATLTILVTDTNDHPVEBQOEYKESLRE 300  
 Db 241 VTTAEELDRETKSTVFRVTAODHGMRRSALATLTILVTDTNDHPVEBQOEYKESLRE 300  
 Qy 301 NLEVEYVLTVAATGDAPPNANILYRLLEGSGSPSEVFEIDPSSGVIKRGKPDREBV 360  
 Db 301 NLEVEYVLTVAATGDAPPNANILYRLLEGSGSPSEVFEIDPSSGVIKRGKPDREBV 360  
 Qy 361 ESYQULVEASDQGRDPGRSTTAAVFLSYEDDNDNAPQSEKRYVQVREEDVTPCAPYLK 420  
 Db 361 ESYQULVEASDQGRDPGRSTTAAVFLSYEDDNDNAPQSEKRYVQVREEDVTPCAPYLK 420  
 Qy 421 VTTASRDVGSNAVHTYSIMSGNARQFYLDAGTALDVVSPLDYETTKBYTLRVAQDGG 480  
 Db 421 VTTASRDVGSNAVHTYSIMSGNARQFYLDAGTALDVVSPLDYETTKBYTLRVAQDGG 480  
 Qy 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESPGLGYLHQAIDADADNARL 540  
 Db 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESPGLGYLHQAIDADADNARL 540  
 Qy 541 EYRLAGVHDEPFTINNGTGMISVAEELDREVDYFSFGVEARDHGTALTSASVSTV 600  
 Db 541 EYRLAGVHDEPFTINNGTGMISVAEELDREVDYFSFGVEARDHGTALTSASVSTV 600  
 Qy 601 LDVNDNPFPOPEYTVRLNEDAVGTSVTVSAVDRDAHSVTVQITSGNTRNFSITS 660  
 Db 601 LDVNDNPFPOPEYTVRLNEDAVGTSVTVSAVDRDAHSVTVQITSGNTRNFSITS 660  
 Qy 661 OSGGGLVSLALPLDYKLEROYVLAATASDGTROPDAQIVNNTADNTRPVPQSSHHTVN 720  
 Db 661 OSGGGLVSLALPLDYKLEROYVLAATASDGTROPDAQIVNNTADNTRPVPQSSHHTVN 720  
 Qy 721 VNEBRPACTTVVLSATDEBTEGNARITYFMEDSIPOFRIDADTGAVTQAELEDVQVS 780  
 Db 721 VNEBRPACTTVVLSATDEBTEGNARITYFMEDSIPOFRIDADTGAVTQAELEDVQVS 780  
 Qy 781 YTLAITARDNGIPQKSDTTTLEILVNDVNDNAPQFLRDSYQSGYEDVPFTSVLQISAT 840  
 Db 781 YTLAITARDNGIPQKSDTTTLEILVNDVNDNAPQFLRDSYQSGYEDVPFTSVLQISAT 840  
 Qy 841 DRDSGLNRFVETPCGDDGDDGDPFVETSGIVRLRLDRBNNAQVTLRAYADKGMF 900  
 Db 841 DRDSGLNRFVETPCGDDGDDGDPFVETSGIVRLRLDRBNNAQVTLRAYADKGMF 900  
 Qy 901 ARTPEVETVTLVDVNDNPFVEQDEDFDVFEENSPIGLAVARVTAATDDEGTNAQIMYQI 960  
 Db 901 ARTPEVETVTLVDVNDNPFVEQDEDFDVFEENSPIGLAVARVTAATDDEGTNAQIMYQI 960  
 Qy 961 VEGNIPEVFOLDIPESGELTALVDLDYEDRPREVIVIQATSAPLVSRATVHRLDRNDP 1020  
 Db 961 VEGNIPEVFOLDIPESGELTALVDLDYEDRPREVIVIQATSAPLVSRATVHRLDRNDP 1020  
 Qy 1021 PVLDNFEELFNNTYTNRSSPFGAIGRVPAHDPDISLSTYSPERGHELSVILNASTG 1080  
 Db 1021 PVLDNFEELFNNTYTNRSSPFGAIGRVPAHDPDISLSTYSPERGHELSVILNASTG 1080  
 Qy 1081 ELKLSRALDNNRPLEAIVSVSDGVHSVTACALRVITITDEMULHSTITRLLEMSPER 1140  
 Db 1081 ELKLSRALDNNRPLEAIVSVSDGVHSVTACALRVITITDEMULHSTITRLLEMSPER 1140  
 Qy 1141 FLSPULGLFIOAVATLTATPPDHVVVNVQDTPDAPGGHIIANVSLVSGQPPGGGPPFL 1200  
 Db 1141 FLSPULGLFIOAVATLTATPPDHVVVNVQDTPDAPGGHIIANVSLVSGQPPGGGPPFL 1200  
 Qy 1201 PSEDLQEBLYNBSLTLTAISAORVLPDPDNTCLAREPCENYRCVSVLRFDSAPPIASSS 1260  
 Db 1201 PSEDLQEBLYNBSLTLTAISAORVLPDPDNTCLAREPCENYRCVSVLRFDSAPPIASSS 1260

Qy 1261 VLFREIHPVGGILRCPCPGFTGDXCETEVDLCYSRPGCPHGRCRBEGGYTCLAGDGYTG 1320  
 Db 1261 VLFREIHPVGGILRCPCPGFTGDXCETEVDLCYSRPGCPHGRCRBEGGYTCLAGDGYTG 1320  
 Qy 1321 EHCEVSASRGCTPGVCNKGTCVNLVGGFKCDCPSGDFEKPYQVYTRSPFPAHSFTTF 1380  
 Db 1321 EHCEVSASRGCTPGVCNKGTCVNLVGGFKCDCPSGDFEKPYQVYTRSPFPAHSFTTF 1380  
 Qy 1381 RGLRORHFPTALSFAATERBDGLLYNGRFNEKIDFVALFVIOBQVLTFSAGSESTTVS 1440  
 Db 1381 RGLRORHFPTALSFAATERBDGLLYNGRFNEKIDFVALFVIOBQVLTFSAGSESTTVS 1440  
 Qy 1441 PFVEGVSVDGQMHVYQLYKTYNKPILLGGTGLPQGSBQVAVVTVYDGCCTGVALARGSYLG 1500  
 Db 1441 PFVEGVSVDGQMHVYQLYKTYNKPILLGGTGLPQGSBQVAVVTVYDGCCTGVALARGSYLG 1500  
 Qy 1501 NYSQAQGTQGSKKSIDLTPGLLLGGVPLPSSFPVARMQFVQCMRLQVDSRHIDMAD 1560  
 Db 1501 NYSQAQGTQGSKKSIDLTPGLLLGGVPLPSSFPVARMQFVQCMRLQVDSRHIDMAD 1560  
 Qy 1561 FIANNGTVPGRPAKKNVCDSTCHNGTCVUNQMAPSCCEPLGFGKSCAQEMANPOHFL 1620  
 Db 1561 FIANNGTVPGRPAKKNVCDSTCHNGTCVUNQMAPSCCEPLGFGKSCAQEMANPOHFL 1620  
 Qy 1621 GSSIVAMHGLSLPIQSPWYLSIMFRTROADGVLLQAITRGSTITTLQREBGMVLSYBGT 1680  
 Db 1621 GSSIVAMHGLSLPIQSPWYLSIMFRTROADGVLLQAITRGSTITTLQREBGMVLSYBGT 1680  
 Qy 1681 GLQASSLRLEPGRANDGDWHHAQALGASGGPGHAILSFYDQGRABGNLGPRLHGLHS 1740  
 Db 1681 GLQASSLRLEPGRANDGDWHHAQALGASGGPGHAILSFYDQGRABGNLGPRLHGLHS 1740  
 Qy 1741 NITVGGIPGPAAGVARGRGLQGVRSVDPREGNSLDPSHGESINTEGGSLDPCDSN 1800  
 Db 1741 NITVGGIPGPAAGVARGRGLQGVRSVDPREGNSLDPSHGESINTEGGSLDPCDSN 1800  
 Qy 1801 PCPANSYCSNDMDPSYSCSPGYGDNCTVCDLNPCEHOSVCRKPSAPRGYTCCECPFN 1860  
 Db 1801 PCPANSYCSNDMDPSYSCSPGYGDNCTVCDLNPCEHOSVCRKPSAPRGYTCCECPFN 1860  
 Qy 1861 YLGPYCESTRIDQCPRGMGWHPITCGPCNCVDSKGFDPDCKNTSGECHKENHYRPPGSPPT 1920  
 Db 1861 YLGPYCESTRIDQCPRGMGWHPITCGPCNCVDSKGFDPDCKNTSGECHKENHYRPPGSPPT 1920  
 Qy 1921 CLLCDYPTGSLSVCDPBDGQCPCKPVGIGRQCDRCNDNPAFVYTTNGCFVNYSCRAI 1980  
 Db 1921 CLLCDYPTGSLSVCDPBDGQCPCKPVGIGRQCDRCNDNPAFVYTTNGCFVNYSCRAI 1980  
 Qy 1981 BAGIWMPTRFGLPAAPCPKGSFGTAVRHCDEHGMPLPVLFACTSITPSELKGFABRL 2040  
 Db 1981 BAGIWMPTRFGLPAAPCPKGSFGTAVRHCDEHGMPLPVLFACTSITPSELKGFABRL 2040  
 Qy 2041 QRNESGLDSGRSQALLLNATQHTAGYFGSDVKAAYQATRLLAHESHTGRGGLSATQ 2100  
 Db 2041 QRNESGLDSGRSQALLLNATQHTAGYFGSDVKAAYQATRLLAHESHTGRGGLSATQ 2100  
 Qy 2101 DVHFTENLTVGSAILLDTANKRHMLIQQTBGTAHLLQYEAASALAQMRRITVLSPF 2160  
 Db 2101 DVHFTENLTVGSAILLDTANKRHMLIQQTBGTAHLLQYEAASALAQMRRITVLSPF 2160  
 Qy 2161 TIVTPNIVISVRLDKNFAAGAKLPRYEALRGEOPPLETTLVILPESVFRETTPVVPAPAG 2220  
 Db 2161 TIVTPNIVISVRLDKNFAAGAKLPRYEALRGEOPPLETTLVILPESVFRETTPVVPAPAG 2220  
 Qy 2221 PGEAOBEBELARRORRHELSQGEAVASVIIYRTLAGLLPHNYDPDRSLRVPKRPPIINT 2280  
 Db 2221 PGEAOBEBELARRORRHELSQGEAVASVIIYRTLAGLLPHNYDPDRSLRVPKRPPIINT 2280  
 Qy 2281 PVSIVSYHDBEELPRALDKPVTVQFRLTERTKRIQVPMNHSIIVSGTGGSGSARGCE 2340  
 Db 2281 PVSIVSYHDBEELPRALDKPVTVQFRLTERTKRIQVPMNHSIIVSGTGGSGSARGCE 2340

QY 2341 VERNESHVSCQCNHMTSPFVLMVDSRRENGEILLPKTLTYVALGVTALALLTPPELTU 2400  
DB 2341 VERNESHVSCQCNHMTSPFVLMVDSRRENGEILLPKTLTYVALGVTALALLTPPELTU 2400  
QY 2401 LRLTRNNGHGRNRLTRALGLAQLVPLLGINODLPACTIVAILLHFTLCTFSWALL 2460  
DB 2401 LRLTRNNGHGRNRLTRALGLAQLVPLLGINODLPACTIVAILLHFTLCTFSWALL 2460  
QY 2461 ALHLALYALTEVDVNTGPMRFTYMLGWVPAFTGLAVGLDPGNGPDCMLSTYDTLI 2520  
DB 2461 ALHLALYALTEVDVNTGPMRFTYMLGWVPAFTGLAVGLDPGNGPDCMLSTYDTLI 2520  
QY 2521 WSPAGVAPAVSVSVFLYLLAARASCAARQGFEGKGPVGLQPSFVALLLSATWLLAL 2580  
DB 2521 WSPAGVAPAVSVSVFLYLLAARASCAARQGFEGKGPVGLQPSFVALLLSATWLLAL 2580  
QY 2581 LSVNSDTLLFHYLPAFCNCIQGFPIPLSYVLSKEVAKALKACSRKPSDPAALTTKSTL 2640  
DB 2581 LSVNSDTLLFHYLPAFCNCIQGFPIPLSYVLSKEVAKALKACSRKPSDPAALTTKSTL 2640  
QY 2641 TSSVNCPSPYADGRLYQPYGDSAGSLHSTSRGSKQPSYIPFLIRBSALNPGQFPGLG 2700  
DB 2641 TSSVNCPSPYADGRLYQPYGDSAGSLHSTSRGSKQPSYIPFLIRBSALNPGQFPGLG 2700  
QY 2701 DPGSLPLEGDDQDHPDSDLSLEDDQSGSYASTHSSDSEEBEERBAAPFGEQG 2760  
DB 2701 DPGSLPLEGDDQDHPDSDLSLEDDQSGSYASTHSSDSEEBEERBAAPFGEQG 2760  
QY 2761 WDSLILPGARLRLHSTPRKQGGPGKAPWPGDGTAKSSSGNGAPBERLRENGALSR 2820  
DB 2761 WDSLILPGARLRLHSTPRKQGGPGKAPWPGDGTAKSSSGNGAPBERLRENGALSR 2820  
QY 2821 EGSIGLPLGSSAAPHKGLIKKCLPTISEKSLRLPLEQCTSSSGSSASBSRGAPPP 2880  
DB 2821 EGSIGLPLGSSAAPHKGLIKKCLPTISEKSLRLPLEQCTSSSGSSASBSRGAPPP 2880  
QY 2881 RPPRQSLQOLGNGVPIAMSIAGTYDEDSGSEPLFFNFLLH 2923  
DB 2881 RPPRQSLQOLGNGVPIAMSIAGTYDEDSGSEPLFFNFLLH 2923

RESULT 4  
US-10-120-801-53  
; Sequence 53, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Shinkete, Richard  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Mehrahan, Fuad  
; APPLICANT: Topper, James N  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT APPLICATION NUMBER: US/10/120, 801  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/285748  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/288334  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/291241

QY PRIOR FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 60/322284  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/285609  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 53  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-120-801-53  
Query Match 100.0%; Score 2923; DB 12; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSBPATGVPPLPPLPPLLLLLLLPPLLDGVPCCSLSGRGSSGACAPMGWLCPS 60  
DB 1 MSBPATGVPPLPPLPPLLLLLLLPPLLDGVPCCSLSGRGSSGACAPMGWLCPS 60  
QY 61 SASNMLVTSRCRDAGTBLTGHVPHHGLRVCPSESAHPLPAPBGCPSGCLLGIG 120  
DB 61 SASNMLVTSRCRDAGTBLTGHVPHHGLRVCPSESAHPLPAPBGCPSGCLLGIG 120  
QY 121 GHLSFGKLTLPBEHPCLKAPRLRCQSCKLAQAPGLRAGERSPEESLGGRRRNVTAPQ 180  
DB 121 GHLSFGKLTLPBEHPCLKAPRLRCQSCKLAQAPGLRAGERSPEESLGGRRRNVTAPQ 180  
QY 181 FQPPSYOATVPENQAGTPVVASLRAIDPDEGAGRLBTMDALPFSRNOFSLDPVICA 240  
DB 181 FQPPSYOATVPENQAGTPVVASLRAIDPDEGAGRLBTMDALPFSRNOFSLDPVICA 240  
QY 241 VTTARELREKTSHTVFVTADHGMPPRSALATLTITVTDNHDPVEQOEYKESLRE 300  
DB 241 VTTARELREKTSHTVFVTADHGMPPRSALATLTITVTDNHDPVEQOEYKESLRE 300  
QY 301 NLEVGVEVLTVRATGDAPAPNANILYRLBGSQSPSEVFEIDPRSGVIRTRGPVREBV 360  
DB 301 NLEVGVEVLTVRATGDAPAPNANILYRLBGSQSPSEVFEIDPRSGVIRTRGPVREBV 360  
QY 361 ESYQUTVERASDQGRPGRRSTTAANVFLSVEDNDNAPQSEKRYVQVREDVTPGAPVLR 420  
DB 361 ESYQUTVERASDQGRPGRRSTTAANVFLSVEDNDNAPQSEKRYVQVREDVTPGAPVLR 420  
QY 421 VTASDRDKSNVAVHYISMSGARQFYLDAGTGLDVVSPLDYETTKETTLRVARADGG 480  
DB 421 VTASDRDKSNVAVHYISMSGARQFYLDAGTGLDVVSPLDYETTKETTLRVARADGG 480  
QY 481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVYLSVPLGYLVHQAIDADAGNARL 540  
DB 481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVYLSVPLGYLVHQAIDADAGNARL 540  
QY 541 EYRLAGVGHDPFTINNGTGMISVAAEIDREVDYVSGVEARHDTGTPALTSASVSTV 600  
DB 541 EYRLAGVGHDPFTINNGTGMISVAAEIDREVDYVSGVEARHDTGTPALTSASVSTV 600  
QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYSAVDRDAHSYITVQITSGNTRNFSITS 660  
DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVYSAVDRDAHSYITVQITSGNTRNFSITS 660  
QY 661 QGGGGLVSLALPLDYKRLERQVYLAVTASDGTQDTAQIVVNTDANTHRPVQSSHYYVN 720  
DB 661 QGGGGLVSLALPLDYKRLERQVYLAVTASDGTQDTAQIVVNTDANTHRPVQSSHYYVN 720  
QY 721 VNEDRPAGTIVVLSATBEDGSENNARITYPMDSIPORPIDDTGAVTQAALDVEDQVS 780  
DB 721 VNEDRPAGTIVVLSATBEDGSENNARITYPMDSIPORPIDDTGAVTQAALDVEDQVS 780  
QY 781 YTLAATTARDNGIPQSDTYYLEILVNDVNDNAPQFLRDSYQGSVYEDVPPFTSVQISAT 840  
DB 781 YTLAATTARDNGIPQSDTYYLEILVNDVNDNAPQFLRDSYQGSVYEDVPPFTSVQISAT 840

QY 841 DRDGLNRRVYTTPOGDDGDDPIVSTSGIVTRLRLDRBNVAQVTLAVADKMP 900  
 DB 841 DRDGLNRRVYTTPOGDDGDDPIVSTSGIVTRLRLDRBNVAQVTLAVADKMP 900  
 QY 901 ARTPEMVTYVTLVDVNDPVPFQDEDFVEENSTIGLAVAVTATDDEGNAIMQI 960  
 DB 901 ARTPEMVTYVTLVDVNDPVPFQDEDFVEENSTIGLAVAVTATDDEGNAIMQI 960  
 QY 961 VEGNIPEVQOLDIFSGELTALVDLDEDRPEVYVLOATSAPLVSRAVTHRLDRNDP 1020  
 DB 961 VEGNIPEVQOLDIFSGELTALVDLDEDRPEVYVLOATSAPLVSRAVTHRLDRNDP 1020  
 QY 1021 PVLGHEFLLFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFPERGMBLSVLINASTG 1080  
 DB 1021 PVLGHEFLLFNNTYNNRSSPPGCAIGVPAHPDISLSLTYSFPERGMBLSVLINASTG 1080  
 QY 1081 ELKLSRALDNNRPLEAIVSVSDGHSVTAQCALRVIIIDEMLTHSITLRLDEMSER 1140  
 DB 1081 ELKLSRALDNNRPLEAIVSVSDGHSVTAQCALRVIIIDEMLTHSITLRLDEMSER 1140  
 QY 1141 FLSPLLGLFIQAVATTLATPDHVVVFNVDTPAPGHHILNVSLSVGQPPGGGPPFL 1200  
 DB 1141 FLSPLLGLFIQAVATTLATPDHVVVFNVDTPAPGHHILNVSLSVGQPPGGGPPFL 1200  
 QY 1201 PSEBQGERLYNRSLLTAISAQRVLPDDNICTAREPCENYRCVSLRFPDSAPFIASS 1260  
 DB 1201 PSEBQGERLYNRSLLTAISAQRVLPDDNICTAREPCENYRCVSLRFPDSAPFIASS 1260  
 QY 1261 VLFPRIHVVGGLRCPCPGFTGDYCTETVDLCYSRPGPHGCRREGGYTCLCDGDTG 1320  
 DB 1261 VLFPRIHVVGGLRCPCPGFTGDYCTETVDLCYSRPGPHGCRREGGYTCLCDGDTG 1320  
 QY 1321 EHCEVASASGRCTPVCNNGGTCVNLVVGFKCPCSDPEKPYCQVTRSPAPHSITP 1380  
 DB 1321 EHCEVASASGRCTPVCNNGGTCVNLVVGFKCPCSDPEKPYCQVTRSPAPHSITP 1380  
 QY 1381 RGLAQRHFHTLALSTATERBGLLYNRRFNEKHPVALLEYTOEVOULTFSAGESITVS 1440  
 DB 1381 RGLAQRHFHTLALSTATERBGLLYNRRFNEKHPVALLEYTOEVOULTFSAGESITVS 1440  
 QY 1441 PFPVPGVSDGOMHTYQLKYNNKPLIGOTGLPQSPESOKAVAVTVVDCCTGTALRGSITG 1500  
 DB 1441 PFPVPGVSDGOMHTYQLKYNNKPLIGOTGLPQSPESOKAVAVTVVDCCTGTALRGSITG 1500  
 QY 1501 NYSCAAQGTGSGSKSLDLTGPELLIGVPLDPESPVAMROFVGCNMLQVDSRIIDAD 1560  
 DB 1501 NYSCAAQGTGSGSKSLDLTGPELLIGVPLDPESPVAMROFVGCNMLQVDSRIIDAD 1560  
 QY 1561 FIANNGTVPCCPAKKNVCDNSTCHNGGTCVNOMDAFSCCEPLGRGSKCAQBMANPOHFL 1620  
 DB 1561 FIANNGTVPCCPAKKNVCDNSTCHNGGTCVNOMDAFSCCEPLGRGSKCAQBMANPOHFL 1620  
 QY 1621 GSSLIYAHMGLSLPISQPMYLSIMFTRQADGVLLOAIRGSTITLQLRBGMVLSVGT 1680  
 DB 1621 GSSLIYAHMGLSLPISQPMYLSIMFTRQADGVLLOAIRGSTITLQLRBGMVLSVGT 1680  
 QY 1681 GLQASSLALEPGRANDGMHAQALGASGGPBAHLSFDYQOQQAEBNLRHGLHLS 1740  
 DB 1681 GLQASSLALEPGRANDGMHAQALGASGGPBAHLSFDYQOQQAEBNLRHGLHLS 1740  
 QY 1741 NITVGGITGPAGVARGGCLQYRVSDTPEGVNSLDPHSGESINVOGCSLDPCCSN 1800  
 DB 1741 NITVGGITGPAGVARGGCLQYRVSDTPEGVNSLDPHSGESINVOGCSLDPCCSN 1800  
 QY 1801 PCPANSYCSNMDSVSCDPCGYGDNCTNVCNLPCEHOSVCTKPSAPHGTYCPCPN 1860  
 DB 1801 PCPANSYCSNMDSVSCDPCGYGDNCTNVCNLPCEHOSVCTKPSAPHGTYCPCPN 1860  
 QY 1861 YLGPYCETRIIDQPCRGWGHPTGCPNCNVSKGFPDCCNTSGECHKENHYRPPSGPT 1920  
 DB 1861 YLGPYCETRIIDQPCRGWGHPTGCPNCNVSKGFPDCCNTSGECHKENHYRPPSGPT 1920  
 QY 1921 CLDCDCYPTGSLSRVCDPEBQCCPKGPIVIGRQCDRCNPNPAEYVTTNCEVNYDSCPAI 1980

DB 1921 CLDCDCYPTGSLSRVCDPEBQCCPKGPIVIGRQCDRCNPNPAEYVTTNCEVNYDSCPAI 1980  
 QY 1981 EAGIWMPTRFGLPAAAPCPKSGFATAVRHCDERHGMPLPULFNCTSTTFSELGFAERL 2040  
 DB 1981 EAGIWMPTRFGLPAAAPCPKSGFATAVRHCDERHGMPLPULFNCTSTTFSELGFAERL 2040  
 QY 2041 QRNESGLDSRSGOQALALRNATQTAGVFGSDVKAVALATRLLAHSTORGFLSATQ 2100  
 DB 2041 QRNESGLDSRSGOQALALRNATQTAGVFGSDVKAVALATRLLAHSTORGFLSATQ 2100  
 QY 2101 DVHFTENILRVGSLALDTANKRHEMLQOTEGGTAMLOHYEAYASALAOHMRHTYLSF 2160  
 DB 2101 DVHFTENILRVGSLALDTANKRHEMLQOTEGGTAMLOHYEAYASALAOHMRHTYLSF 2160  
 QY 2161 TIVTENVISVVRDKNPFAGAKLPRYEAALRGEPDLETTVIIPESVFRETPIVVRPAG 2220  
 DB 2161 TIVTENVISVVRDKNPFAGAKLPRYEAALRGEPDLETTVIIPESVFRETPIVVRPAG 2220  
 QY 2221 PGEAQBEBELARRQRHPELSQGEAVASVITRTIAGLLPNNDPDKRSLRVPKPIINT 2280  
 DB 2221 PGEAQBEBELARRQRHPELSQGEAVASVITRTIAGLLPNNDPDKRSLRVPKPIINT 2280  
 QY 2281 PVVSIYVHDBELLPRALDRVTVQFRLTEERTKPICVFMNHSILVSGTGSASGCE 2340  
 DB 2281 PVVSIYVHDBELLPRALDRVTVQFRLTEERTKPICVFMNHSILVSGTGSASGCE 2340  
 QY 2341 VVFNESHVSQCNHMTSPAVLMDVSRRENGEILPLTLTVAGLVTLAALLTFPEFLT 2400  
 DB 2341 VVFNESHVSQCNHMTSPAVLMDVSRRENGEILPLTLTVAGLVTLAALLTFPEFLT 2400  
 QY 2401 LRILRSNOHIGIRNLTALGLAQVFLGGINQADLPACTVIALHFLYCTESMALLE 2460  
 DB 2401 LRILRSNOHIGIRNLTALGLAQVFLGGINQADLPACTVIALHFLYCTESMALLE 2460  
 QY 2461 ALHLRYALTEVRDVTNCPMRFYTLGNGVPAFTIGLAVGDPBEYGNPDGMLSYTTLI 2520  
 DB 2461 ALHLRYALTEVRDVTNCPMRFYTLGNGVPAFTIGLAVGDPBEYGNPDGMLSYTTLI 2520  
 QY 2521 WSPAGPVAFAVMSVFLYITLAARASCAORGFEEKKPFVSGLOPSFVLLLSATWLLAL 2580  
 DB 2521 WSPAGPVAFAVMSVFLYITLAARASCAORGFEEKKPFVSGLOPSFVLLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHYLPATNCIOGPEIFLSYVLSKVRKALKACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHYLPATNCIOGPEIFLSYVLSKVRKALKACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSVNCSPYADGRLYQPYGDSAGSLHSTRSGKSPSYTPTLLREBSALNPGGPPGLG 2700  
 DB 2641 TSSVNCSPYADGRLYQPYGDSAGSLHSTRSGKSPSYTPTLLREBSALNPGGPPGLG 2700  
 QY 2701 DPGSLFLEGDDQDHPDTSDDSLSDPDQSGSYASTHSSDSBEEBEEBEAFAPEOG 2760  
 DB 2701 DPGSLFLEGDDQDHPDTSDDSLSDPDQSGSYASTHSSDSBEEBEEBEAFAPEOG 2760  
 QY 2761 WDSLIGFAERLPLHSTPKDGGPGPKAPWPGDFTAKSSGNGAPBERRLENGALSR 2820  
 DB 2761 WDSLIGFAERLPLHSTPKDGGPGPKAPWPGDFTAKSSGNGAPBERRLENGALSR 2820  
 QY 2821 EGSIGPLPGSSAOPHKIILKKKCLPTISSEKSLRLPLEOCTGSSRSKSSASBEGSGGPP 2880  
 DB 2821 EGSIGPLPGSSAOPHKIILKKKCLPTISSEKSLRLPLEOCTGSSRSKSSASBEGSGGPP 2880  
 QY 2881 RPPRQSLQOLNGVMPFIAMSIRKATVDEBSSGSEFLPFNFILH 2923  
 DB 2881 RPPRQSLQOLNGVMPFIAMSIRKATVDEBSSGSEFLPFNFILH 2923

RESULT 5  
 us-10-292-798-932  
 ; Sequence 932, Application us/10292798  
 ; Publication No. US2003023583A1  
 ; GENERAL INFORMATION:

APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 932  
LENGTH: 2923  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-292-798-932

Query Match 100.0%; Score 2923; DB 12; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESPATGVLPPTPPPLLLLLLLLLLPPPLADQVGCPSLSGRGSSGACAPMGMLCPSS 60  
DB 1 MESPATGVLPPTPPPLLLLLLLLLLPPPLADQVGCPSLSGRGSSGACAPMGMLCPSS 60

QY 61 SASNLMLYTSRCDAGTELTGHLVPHHDLRVWCPSSEAHIPLPAPAECCPMSCRLGIG 120  
DB 61 SASNLMLYTSRCDAGTELTGHLVPHHDLRVWCPSSEAHIPLPAPAECCPMSCRLGIG 120

QY 121 GHLSPOGKLTLPBEPCLAPRLRCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180  
DB 121 GHLSPOGKLTLPBEPCLAPRLRCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180

QY 121 GHLSPOGKLTLPBEPCLAPRLRCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180  
DB 121 GHLSPOGKLTLPBEPCLAPRLRCOSCKLAQAPGLRAGERSPEESLGGRRRRNVTAQ 180

QY 181 POPPSQATVPENOPAGTVAISRAIDPDEGEAGLEETMDALFDSRSNOFSLDPVTA 240  
DB 181 POPPSQATVPENOPAGTVAISRAIDPDEGEAGLEETMDALFDSRSNOFSLDPVTA 240

QY 241 VTTAEELDEKSTSTHYFRVTAODHGMRRSALATLTILVTDNHDHPVEQOEYKESLRE 300  
DB 241 VTTAEELDEKSTSTHYFRVTAODHGMRRSALATLTILVTDNHDHPVEQOEYKESLRE 300

QY 301 NIEVGYEVLTVAITDGDAPPNANILYRLIEGSGSSSEVEIDPRSGVIRTRGVDRREY 360  
DB 301 NIEVGYEVLTVAITDGDAPPNANILYRLIEGSGSSSEVEIDPRSGVIRTRGVDRREY 360

QY 361 ESYVLVEASDQGRDGPSTTAAVFLSVEDDNDNAPOSEKRYVQVREEDVTPGAIVLR 420  
DB 361 ESYVLVEASDQGRDGPSTTAAVFLSVEDDNDNAPOSEKRYVQVREEDVTPGAIVLR 420

QY 421 VTASDRKSNVAVHSIMSGNARQFYLDAGTALDVVSPLDYETTKETTLRVARQDGG 480  
DB 421 VTASDRKSNVAVHSIMSGNARQFYLDAGTALDVVSPLDYETTKETTLRVARQDGG 480

QY 481 RPPLSVSGLVTVQVLDINDNAPIFVSTPQATVLESVPLGYLVTHVQALDADAGNARL 540  
DB 481 RPPLSVSGLVTVQVLDINDNAPIFVSTPQATVLESVPLGYLVTHVQALDADAGNARL 540

QY 541 EYRLGVGHPPTINGTGMISVAEELDBREVDFTSPVTEARDHCTPALTSASVSTV 600  
DB 541 EYRLGVGHPPTINGTGMISVAEELDBREVDFTSPVTEARDHCTPALTSASVSTV 600

QY 601 LDVNDNPTFTQPEYTVRLNEDAAVGTSTVTVSAVDRDAHSVITYOITSQNTNRPSITS 660  
DB 601 LDVNDNPTFTQPEYTVRLNEDAAVGTSTVTVSAVDRDAHSVITYOITSQNTNRPSITS 660

QY 661 QSGGGVSLALPLDYTLERQYVLAATASGTRODTQIYVNTDANTHRPVFQSSHYTN 720  
DB 661 QSGGGVSLALPLDYTLERQYVLAATASGTRODTQIYVNTDANTHRPVFQSSHYTN 720

QY 721 VNEBRPAGTTVVLSATDEDTGENARITYFMEDSIQPRIDADTGAVTTQAEILDYEDQVS 780

DB 721 VNEBRPAGTTVVLSATDEDTGENARITYFMEDSIQPRIDADTGAVTTQAEILDYEDQVS 780

QY 761 YTLAATDANGIPKSDTTYEILVNDVNDNAPOGLRDSYOGSYVEDVPFVSVOISAT 840  
DB 761 YTLAATDANGIPKSDTTYEILVNDVNDNAPOGLRDSYOGSYVEDVPFVSVOISAT 840

QY 841 DRDSGLNGRVEYTFQGGDGDGDFIVESTSGIVTLRLDRENNVAQYVLRAYADKGNP 900  
DB 841 DRDSGLNGRVEYTFQGGDGDGDFIVESTSGIVTLRLDRENNVAQYVLRAYADKGNP 900

QY 901 ARTPEVTVVLDVNDNPPVREODRPFVRENSPTGLAVARVATDDEGNAIMQI 960  
DB 901 ARTPEVTVVLDVNDNPPVREODRPFVRENSPTGLAVARVATDDEGNAIMQI 960

QY 961 VEGNIPFVOLDIPSGELTALVDLDYEDRPREVTVIOATSAPLVSRAVTHVRLDRNDP 1020  
DB 961 VEGNIPFVOLDIPSGELTALVDLDYEDRPREVTVIOATSAPLVSRAVTHVRLDRNDP 1020

QY 1021 PVLGNFEILFNNTYNNRSSFPFGAIGRPAPHPDISDLYTSPERGNELSLVLLNASTG 1080  
DB 1021 PVLGNFEILFNNTYNNRSSFPFGAIGRPAPHPDISDLYTSPERGNELSLVLLNASTG 1080

QY 1081 ELKLSRALDNNRPLEATMSVTVSDGVSHTAQCALRTVITIDEMLTHSITTLRLDMSR 1140  
DB 1081 ELKLSRALDNNRPLEATMSVTVSDGVSHTAQCALRTVITIDEMLTHSITTLRLDMSR 1140

QY 1141 FLSPLLGLFIQAVAAATLATPPDHVVVFNQDRTPAPEGHILNVSLVQCPGPGGPPFL 1200  
DB 1141 FLSPLLGLFIQAVAAATLATPPDHVVVFNQDRTPAPEGHILNVSLVQCPGPGGPPFL 1200

QY 1201 PSEDIQERLYTNRSILTAISAQRYLPFDNTCLREPCENYRCVSLRPSSAPFIASS 1260  
DB 1201 PSEDIQERLYTNRSILTAISAQRYLPFDNTCLREPCENYRCVSLRPSSAPFIASS 1260

QY 1261 VLFRPIHVGGLRCPCPFTGDYCEFEVDLCYSRPGCPHRCRSREGYCLCDGVTG 1320  
DB 1261 VLFRPIHVGGLRCPCPFTGDYCEFEVDLCYSRPGCPHRCRSREGYCLCDGVTG 1320

QY 1321 EHCEVSARSGRCTPVCNKGTCVNLVVGCFKDCPSPGFEKPYCQVTRSFPAHSFTTF 1380  
DB 1321 EHCEVSARSGRCTPVCNKGTCVNLVVGCFKDCPSPGFEKPYCQVTRSFPAHSFTTF 1380

QY 1381 RGLRQRFHTLALSPATFERDGLLYNGRFNEKDFVALBYIOROVLTFSSAGSSTTVS 1440  
DB 1381 RGLRQRFHTLALSPATFERDGLLYNGRFNEKDFVALBYIOROVLTFSSAGSSTTVS 1440

QY 1441 PVPBGVSDGQWHYVQLKRYNKPILGOTGLPOGSEOKVAIVTVDGCDTGVALARGSYLG 1500  
DB 1441 PVPBGVSDGQWHYVQLKRYNKPILGOTGLPOGSEOKVAIVTVDGCDTGVALARGSYLG 1500

QY 1501 NVSCAAGTQGGSKKSLDLTGPLLGGVPDLPESEFPVAMROFVGCGRNLOVDSRIIDNAD 1560  
DB 1501 NVSCAAGTQGGSKKSLDLTGPLLGGVPDLPESEFPVAMROFVGCGRNLOVDSRIIDNAD 1560

QY 1561 FIANNGTVPCCPAKKNVCDNSTCANGTCTVNQMDAFSCBCLGFGGKSCAOMANPOHFL 1620  
DB 1561 FIANNGTVPCCPAKKNVCDNSTCANGTCTVNQMDAFSCBCLGFGGKSCAOMANPOHFL 1620

QY 1621 GSSLVAMHGLSLPSOPWYLSLMPRTROADVLLQAIRGSGSTTLQIRBGVMSVSGT 1680  
DB 1621 GSSLVAMHGLSLPSOPWYLSLMPRTROADVLLQAIRGSGSTTLQIRBGVMSVSGT 1680

QY 1681 GLQASLRLBGRANDGDMHAQALASAGGPGHAIIISFDYGOQRAEGNLPRLHGLHS 1740  
DB 1681 GLQASLRLBGRANDGDMHAQALASAGGPGHAIIISFDYGOQRAEGNLPRLHGLHS 1740

QY 1741 NITVGGIIGPAGVARGRGCLQGYRVSDTREGNLSLPSHGESINVOGCSLPDPCSN 1800  
DB 1741 NITVGGIIGPAGVARGRGCLQGYRVSDTREGNLSLPSHGESINVOGCSLPDPCSN 1800

QY 1801 PCPANSYCSNDMSYSCSDGPGYGDNCTNVCIDNPCEHOSVCTRKPAPAGYTCCECPN 1860



```

Db      1801 PCPANSYCSNDMSYSCSDPGYTGNDCTVNCIDLPNCHEOSVCTKRPAPDHGTCCECPN 1860
Qy      1861 YLGPYCETRIIDQCPBGMWGHPTGCPNCNDVSKGPPDCKNTSGECHKENHRRPPSPY 1920
Db      1861 YLGPYCETRIIDQCPBGMWGHPTGCPNCNDVSKGPPDCKNTSGECHKENHRRPPSPY 1920
Qy      1921 CILCDQCPYSGSLRYCDPEDGQCPCKPQYIGROCDRCNDPFAVTTNGCEVNTDSCPRAI 1980
Db      1921 CILCDQCPYSGSLRYCDPEDGQCPCKPQYIGROCDRCNDPFAVTTNGCEVNTDSCPRAI 1980
Qy      1981 BAGIMWPRTRFGLPALAACPCKSGFGTAVRHCDHRGMLPNLNPCTSIIPSELKGAERL 2040
Db      1981 BAGIMWPRTRFGLPALAACPCKSGFGTAVRHCDHRGMLPNLNPCTSIIPSELKGAERL 2040
Qy      2041 QNRESGLDGRSQQLALLRNATQHTAGYFGSDVKAAYQALATRLAHSTORGFGLSAQ 2100
Db      2041 QNRESGLDGRSQQLALLRNATQHTAGYFGSDVKAAYQALATRLAHSTORGFGLSAQ 2100
Qy      2101 DVHFTENLRVGSALLDTANKHWEILQOTEGGTAMLQHTAAYASALAQNMRHTYLSPP 2160
Db      2101 DVHFTENLRVGSALLDTANKHWEILQOTEGGTAMLQHTAAYASALAQNMRHTYLSPP 2160
Qy      2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEPDLFTTYILPESVREHPVVRPAG 2220
Db      2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEPDLFTTYILPESVREHPVVRPAG 2220
Qy      2221 PGEAOEPEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLRYPKPIINT 2280
Db      2221 PGEAOEPEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLRYPKPIINT 2280
Qy      2281 PIVVSIYVHDEBELPRALDKPVTVOFRLLETERTPICVFNHSLIVSGTGMSARCE 2340
Db      2281 PIVVSIYVHDEBELPRALDKPVTVOFRLLETERTPICVFNHSLIVSGTGMSARCE 2340
Qy      2341 VVFRRESHVSQCQNMHTSPAVLMDVSRRENGELPLKLTLYVALAGTVALALLTTPFELT 2400
Db      2341 VVFRRESHVSQCQNMHTSPAVLMDVSRRENGELPLKLTLYVALAGTVALALLTTPFELT 2400
Qy      2401 IRLISNQHIGIRRNITAAALGAQVFLGINOADLPFACTVIALILHLYLCTFSWALIE 2460
Db      2401 IRLISNQHIGIRRNITAAALGAQVFLGINOADLPFACTVIALILHLYLCTFSWALIE 2460
Qy      2461 ALHLRYALTEVDVNTGPMRFYTMGMGPAPITGLAVGLDPBGKGNPFCWLSTYDTLI 2520
Db      2461 ALHLRYALTEVDVNTGPMRFYTMGMGPAPITGLAVGLDPBGKGNPFCWLSTYDTLI 2520
Qy      2521 WSPAPVAVAVMSVFLYILTAARASCAORQFBEKKGPVSGLOPSPAVLILLSATWILAL 2580
Db      2521 WSPAPVAVAVMSVFLYILTAARASCAORQFBEKKGPVSGLOPSPAVLILLSATWILAL 2580
Qy      2581 LSVNSDTLLFHYLPAFCNCIQQPPIFLASYVVLKSKVRKALKLACSRKPSDPALTTKSTL 2640
Db      2581 LSVNSDTLLFHYLPAFCNCIQQPPIFLASYVVLKSKVRKALKLACSRKPSDPALTTKSTL 2640
Qy      2641 TSSNYCPSPYAAGRLYOPFGDSAGSLHSTRSGKQSPSTIPLLBESALNAGQCPGIG 2700
Db      2641 TSSNYCPSPYAAGRLYOPFGDSAGSLHSTRSGKQSPSTIPLLBESALNAGQCPGIG 2700
Qy      2701 DPGSLFLEGOQOHPDSTDSDLSLEDQSGSYASTHSDBEEREBEEREAAPRGEG 2760
Db      2701 DPGSLFLEGOQOHPDSTDSDLSLEDQSGSYASTHSDBEEREBEEREAAPRGEG 2760
Qy      2761 WDSLLGPGAEURLPLHSTPCKDGPFGKAPWPDFTTAKSSGNGAPBEERLRENDALSR 2820
Db      2761 WDSLLGPGAEURLPLHSTPCKDGPFGKAPWPDFTTAKSSGNGAPBEERLRENDALSR 2820
Qy      2821 BBSLGPPLGSSAOPHKGILKKKCLPTISKSSLRLPLEQCTGSSRGSSASGSGGPP 2880
Db      2821 BBSLGPPLGSSAOPHKGILKKKCLPTISKSSLRLPLEQCTGSSRGSSASGSGGPP 2880
Qy      2881 RPPRQSLQEQNLGVPILMSIKAGTVDEDSGSEFLFNFPLH 2923
Db      2881 RPPRQSLQEQNLGVPILMSIKAGTVDEDSGSEFLFNFPLH 2923

```

```

RESULT 6
US-10-225-567A-524
; Sequence 524, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glenn C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 524
; LENGTH: 2923
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-524

```

```

Query Match      100.0%; Score 2923; DB 15; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MSPATGVPPLPPEPPLLLLLLLLLLPPPLGDQVGPCKSLGSGGSGGACAPMGLCP 60
Db      1 MSPATGVPPLPPEPPLLLLLLLLLLPPPLGDQVGPCKSLGSGGSGGACAPMGLCP 60
Qy      61 SASNMLYTSRCRDAGTELTHIVPHDGLRVWCESRAHILPLPAPGCPMGLGIG 120
Db      61 SASNMLYTSRCRDAGTELTHIVPHDGLRVWCESRAHILPLPAPGCPMGLGIG 120
Qy      121 GHLSPOGRLTPEBHPCLAKAPRLRCQSCKLAQAPGLARAGERSPEBSLGRKRANVTAPQ 180
Db      121 GHLSPOGRLTPEBHPCLAKAPRLRCQSCKLAQAPGLARAGERSPEBSLGRKRANVTAPQ 180
Qy      181 FQPSYQATVBNOPAGTPVASLRAIDPBGAGRLTYMDALPDRSNQFSLDPVTGA 240
Db      181 FQPSYQATVBNOPAGTPVASLRAIDPBGAGRLTYMDALPDRSNQFSLDPVTGA 240
Qy      241 VTTABELRETKSTHVFVTAQDHGMPRRSALATLTILVTQNDHPVFBQOEYKESLRE 300
Db      241 VTTABELRETKSTHVFVTAQDHGMPRRSALATLTILVTQNDHPVFBQOEYKESLRE 300
Qy      301 NLEVGVEVLTVRAATDGDAPPNANILYRLBSGSGSPSEVFEDIPRSGVIRTRGPVDRREV 360
Db      301 NLEVGVEVLTVRAATDGDAPPNANILYRLBSGSGSPSEVFEDIPRSGVIRTRGPVDRREV 360
Qy      361 ESYQLTVAASQGDGPRGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREDEVTPGAPVLR 420
Db      361 ESYQLTVAASQGDGPRGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREDEVTPGAPVLR 420
Qy      421 VTASDRDGSNAVHYHISMSGNARQGYFLDQGTALDVVSFLDVEETTEYTLRVAQDGG 480
Db      421 VTASDRDGSNAVHYHISMSGNARQGYFLDQGTALDVVSFLDVEETTEYTLRVAQDGG 480
Qy      481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVHLVQAIDADGDNARL 540
Db      481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVHLVQAIDADGDNARL 540
Qy      541 EYRLAGVGHDPPTINNGTGWISVAABLDREEDVPSYSGVABRHHGTALTAASASVSTV 600
Db      541 EYRLAGVGHDPPTINNGTGWISVAABLDREEDVPSYSGVABRHHGTALTAASASVSTV 600
Qy      601 LDVVDNNPTFQPEYTVRLNEDAAVGTSVVVSADVDRASHVITVQISGTVRRRFSITS 660
Db      601 LDVVDNNPTFQPEYTVRLNEDAAVGTSVVVSADVDRASHVITVQISGTVRRRFSITS 660

```



QY 661 QSGGGIVSLALPLDYLERQYVLAATASGTRDTAQIVNNTDANTHRPVFOSSHTYN 720  
 DB 661 QSGGGIVSLALPLDYLERQYVLAATASGTRDTAQIVNNTDANTHRPVFOSSHTYN 720  
 QY 721 VNEBRAGTIVYLAISDEDTGENARITYFMEDSIQFRIADDTGAVTQALDYDOYS 780  
 DB 721 VNEBRAGTIVYLAISDEDTGENARITYFMEDSIQFRIADDTGAVTQALDYDOYS 780  
 QY 781 YTLATARDNGIPQKSDITYLEILVNDVNDNAPOFLRDSYQGSVEDVPEPTSVLQISAT 840  
 DB 781 YTLATARDNGIPQKSDITYLEILVNDVNDNAPOFLRDSYQGSVEDVPEPTSVLQISAT 840  
 QY 841 DSDSGANGVFTYFOGDDGDEDFIYESISGIVRTLRDLRENAOYVLAAYDXKMP 900  
 DB 841 DSDSGANGVFTYFOGDDGDEDFIYESISGIVRTLRDLRENAOYVLAAYDXKMP 900  
 QY 901 APTPEVTVTVLVNDNPPVPEODEPDAVEENSPIGLAVARTANDPBGNAOIMYOI 960  
 DB 901 APTPEVTVTVLVNDNPPVPEODEPDAVEENSPIGLAVARTANDPBGNAOIMYOI 960  
 QY 961 VEGNIPEVFOIDIFSGELTALVDLYEDRPEVYVLOATSAPLVSRATVHRLDRNDP 1020  
 DB 961 VEGNIPEVFOIDIFSGELTALVDLYEDRPEVYVLOATSAPLVSRATVHRLDRNDP 1020  
 QY 1021 PVLGNPEILPNNYVTRSSSPFGGAIGRVPAHDPDISLSITTSFERGELSLVLNASTG 1080  
 DB 1021 PVLGNPEILPNNYVTRSSSPFGGAIGRVPAHDPDISLSITTSFERGELSLVLNASTG 1080  
 QY 1081 ELKLSALNNRPLEAIMSVLSDGVHASTAOCALVTIITDEMLHSTLRLRDMSPR 1140  
 DB 1081 ELKLSALNNRPLEAIMSVLSDGVHASTAOCALVTIITDEMLHSTLRLRDMSPR 1140  
 QY 1141 FLSPLIGLFIQAAVATLATPPDHVVFNVQRTDABGHI LNVSLVGQPPGSGGPPPL 1200  
 DB 1141 FLSPLIGLFIQAAVATLATPPDHVVFNVQRTDABGHI LNVSLVGQPPGSGGPPPL 1200  
 QY 1201 PEBDQERLYNRSLTLTASAOVLPFDNICLRECEYMRVSVLARDSSAPFLASS 1260  
 DB 1201 PEBDQERLYNRSLTLTASAOVLPFDNICLRECEYMRVSVLARDSSAPFLASS 1260  
 QY 1261 VLFPRPIHPVGLRCRCPFGTGDYCTEVDLCYSRCGPHGRCSREGGTCICRDGYG 1320  
 DB 1261 VLFPRPIHPVGLRCRCPFGTGDYCTEVDLCYSRCGPHGRCSREGGTCICRDGYG 1320  
 QY 1321 EHCEVASRSGRCTPGVCKNGGTCVNLVGGFKCDPSGDPFKERYCQVTRSPASHFITF 1380  
 DB 1321 EHCEVASRSGRCTPGVCKNGGTCVNLVGGFKCDPSGDPFKERYCQVTRSPASHFITF 1380  
 QY 1381 RGLRORFHFTLALSFAKXERDGLLYNGRPNKXDPVALVIOBOVQLTFSAGBSTTYS 1440  
 DB 1381 RGLRORFHFTLALSFAKXERDGLLYNGRPNKXDPVALVIOBOVQLTFSAGBSTTYS 1440  
 QY 1441 PVPBGVSGOMHTVLOKTYNKPLLGOTGLPOGSPBQKAVAVVTVDGCDGVALRFGSVLG 1500  
 DB 1441 PVPBGVSGOMHTVLOKTYNKPLLGOTGLPOGSPBQKAVAVVTVDGCDGVALRFGSVLG 1500  
 QY 1501 NYSCAAQGTQSGSKSLDTGLPLLLGAVPDLPESPFVRMRQFVGCNRNLQVDSRIHDMAD 1560  
 DB 1501 NYSCAAQGTQSGSKSLDTGLPLLLGAVPDLPESPFVRMRQFVGCNRNLQVDSRIHDMAD 1560  
 QY 1561 FTANNNTVGCPCPKKRVCSNNTCHNGTCVNMWDASCECPILGFGKSCAQEAMANOHL 1620  
 DB 1561 FTANNNTVGCPCPKKRVCSNNTCHNGTCVNMWDASCECPILGFGKSCAQEAMANOHL 1620  
 QY 1621 GSSLVAMHGLSLPISOPWYLSLMTFRTOADGVLLQAITRGRSTITTLQLEBHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISOPWYLSLMTFRTOADGVLLQAITRGRSTITTLQLEBHVMSVEGT 1680  
 QY 1681 GLQASSILRLEPRGANDGWHALQALGASGPGHALISPYGQORAKNGIAPPLHGLHLS 1740  
 DB 1681 GLQASSILRLEPRGANDGWHALQALGASGPGHALISPYGQORAKNGIAPPLHGLHLS 1740  
 QY 1741 NITVGGIPBAGVAGFRGCTLGVRVSDTPBGSVNSLDPHGESINVEQCSLPDPCSN 1800

DB 1741 NITVGGIPBAGVAGFRGCTLGVRVSDTPBGSVNSLDPHGESINVEQCSLPDPCSN 1800  
 QY 1801 PCPANSYCSNMDSDSCSCDPQYGDNCNVCDLNPECHEOSCTKXPSAPHGTCPCPN 1860  
 DB 1801 PCPANSYCSNMDSDSCSCDPQYGDNCNVCDLNPECHEOSCTKXPSAPHGTCPCPN 1860  
 QY 1861 YLGPYCETRIDQPCBGMWGHPTGCPNCNCDVSKGFPDPCNKTSGECHKENHYPGSPPT 1920  
 DB 1861 YLGPYCETRIDQPCBGMWGHPTGCPNCNCDVSKGFPDPCNKTSGECHKENHYPGSPPT 1920  
 QY 1921 CLICDCYTGSLSRVCDPEDDQCPCKRGVIGRQCCRCNPNPABVTYNNCEVNYDSCPAI 1980  
 DB 1921 CLICDCYTGSLSRVCDPEDDQCPCKRGVIGRQCCRCNPNPABVTYNNCEVNYDSCPAI 1980  
 QY 1981 BAGIWMPTRRFGLPAAACPKPSFGTAVRHCDERHGMPLPULFNCTSIITFSELKOPARL 2040  
 DB 1981 BAGIWMPTRRFGLPAAACPKPSFGTAVRHCDERHGMPLPULFNCTSIITFSELKOPARL 2040  
 QY 2041 QNBSGLDSGSSQOALALLRNATOHTAGYFGSDVAVAYOLATRLAHSTORGFLSATQ 2100  
 DB 2041 QNBSGLDSGSSQOALALLRNATOHTAGYFGSDVAVAYOLATRLAHSTORGFLSATQ 2100  
 QY 2101 DVHFTENILRVGSALLDPANRRHWEILQOTEGTAMLLQHEAVASALAOQNNRHITYLSP 2160  
 DB 2101 DVHFTENILRVGSALLDPANRRHWEILQOTEGTAMLLQHEAVASALAOQNNRHITYLSP 2160  
 QY 2161 TIVTNIYISVVRLDKGNFAGAKLPRVYALRGQPPDLETTYILPESVFERETPPVVRBAG 2220  
 DB 2161 TIVTNIYISVVRLDKGNFAGAKLPRVYALRGQPPDLETTYILPESVFERETPPVVRBAG 2220  
 QY 2221 PGEAOPEBELARORRHPLSQGEAVASVITYRTLAGLLPHNYDDPKSLRVPKPIINT 2280  
 DB 2221 PGEAOPEBELARORRHPLSQGEAVASVITYRTLAGLLPHNYDDPKSLRVPKPIINT 2280  
 QY 2281 PVVSIYVHDEBELPRALDKPTVOFRLJTBERTKPICVPMNHSILVSGGMSARGOE 2340  
 DB 2281 PVVSIYVHDEBELPRALDKPTVOFRLJTBERTKPICVPMNHSILVSGGMSARGOE 2340  
 QY 2341 VVFRNESHVSCCNHMTSPAVLMDVSRRENGIILPKLTJYVALGVTLLAALLTFEFTL 2400  
 DB 2341 VVFRNESHVSCCNHMTSPAVLMDVSRRENGIILPKLTJYVALGVTLLAALLTFEFTL 2400  
 QY 2401 LRIILSNOHGIRNUTALGLAQLVFLGINOADI PACCTYAIILHFLYCTEFWALLE 2460  
 DB 2401 LRIILSNOHGIRNUTALGLAQLVFLGINOADI PACCTYAIILHFLYCTEFWALLE 2460  
 QY 2461 ALHLRYALTBVRDVTNGMRFYMLGMVPAFITGLAVGLDBEGYGNDFCMLSITYDLI 2520  
 DB 2461 ALHLRYALTBVRDVTNGMRFYMLGMVPAFITGLAVGLDBEGYGNDFCMLSITYDLI 2520  
 QY 2521 WSFAPVAVAVMSVFLYTLAARASCAAOQOFKXGVSGLQPSFAVILLLSATWLLAL 2580  
 DB 2521 WSFAPVAVAVMSVFLYTLAARASCAAOQOFKXGVSGLQPSFAVILLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLPATCNCIGPPIFLSYVLSKXVRKALKLACGRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLPATCNCIGPPIFLSYVLSKXVRKALKLACGRKSPDPALTTKSTL 2640  
 QY 2641 TSSVNCSPYADGRLYOPYGSAGSLHSTSRGSKQPSYIPTLAEBSALNPGQPPGIG 2700  
 DB 2641 TSSVNCSPYADGRLYOPYGSAGSLHSTSRGSKQPSYIPTLAEBSALNPGQPPGIG 2700  
 QY 2701 DFGSLFLBGQDQHDPTDSDSLLEDDQGSVASTHSDBEEEBEERAAEPGEQ 2760  
 DB 2701 DFGSLFLBGQDQHDPTDSDSLLEDDQGSVASTHSDBEEEBEERAAEPGEQ 2760  
 QY 2761 WSLILGPABERLPLHSTKQGGPGPKAPWPCDPCTTKAKESGNGAPREERLENDALSR 2820  
 DB 2761 WSLILGPABERLPLHSTKQGGPGPKAPWPCDPCTTKAKESGNGAPREERLENDALSR 2820  
 QY 2821 ESSLGLPGSSAQPKGILLKKCLPTISEKSLJLPLQCTGSGSGSASGSGSGGPP 2880

Db	2821	EGSLGPLESSAQPPIKGIKKKCLPTISKSLMLPLEQCTGSSRGSSASESGSGPP	2880
Qy	2881	RPPRQSLQEQJNGVPIAMSIKAGTVDESSGSEFLPFNFJH	2923
Db	2881	RPPRQSLQEQJNGVPIAMSIKAGTVDESSGSEFLPFNFJH	2923
RESULT 7			
US-09-788-711A-2			
; Sequence 2, Application US/09788711A			
; Patent No. US2002058328A1			
; GENERAL INFORMATION:			
; APPLICANT: Tania Tamsin Testa			
; TITLE OF INVENTION: NOVEL COMPOUNDS			
; FILE REFERENCE: GP-30225			
; CURRENT APPLICATION NUMBER: US/09/788,711A			
; CURRENT FILING DATE: 2001-02-20			
; PRIOR APPLICATION NUMBER: 0004196.2			
; PRIOR FILING DATE: 2000-02-19			
; NUMBER OF SEQ ID NOS: 4			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 2			
; LENGTH: 2956			
; TYPE: PRT			
; ORGANISM: HOMO SAPIENS			
US-09-788-711A-2			
Query Match 97.1%; Score 2837; DB 9; Length 2956;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MSRATGVPPLTPPPPLLLLLLLLLLPPPLLDGVPCCSLGSRGSSGACAPMGLCP	60
Db	1	MSRATGVPPLTPPPPLLLLLLLLLLPPPLLDGVPCCSLGSRGSSGACAPMGLCP	60
Qy	61	SASNLMLTSCRGRDGTGLTGLVPHHGLRWCESEAHITPLPAPRGCCPMSCLLIG	120
Db	61	SASNLMLTSCRGRDGTGLTGLVPHHGLRWCESEAHITPLPAPRGCCPMSCLLIG	120
Qy	121	GHLSPQGLTLPEEHPCLAKAPRLRCQSCLAQAPGLRAGERSPEBSLGGRRKRVNTAPQ	180
Db	121	GHLSPQGLTLPEEHPCLAKAPRLRCQSCLAQAPGLRAGERSPEBSLGGRRKRVNTAPQ	180
Qy	181	POPSPQATVPENOPAGTPVASLRAIDPDEGAGLETTMDLPDSRNOPESLDPVGA	240
Db	181	POPSPQATVPENOPAGTPVASLRAIDPDEGAGLETTMDLPDSRNOPESLDPVGA	240
Qy	241	VTTAEELRETKSTHVFVTAQDHGMPRRSALA.TITLVTDPNDHPVPEOEYKESLRE	300
Db	241	VTTAEELRETKSTHVFVTAQDHGMPRRSALA.TITLVTDPNDHPVPEOEYKESLRE	300
Qy	301	NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFELIDPSGVIKTRGPDREEV	360
Db	301	NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFELIDPSGVIKTRGPDREEV	360
Qy	361	ESYQLTVEASQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVVOYREDEVTPCAPLR	420
Db	361	ESYQLTVEASQGRDPGRSTTAAVFLSVEDDNDNAPOFSEKRYVVOYREDEVTPCAPLR	420
Qy	421	VTAASRDGSAVVVYYSIMSGNARGOFTLDAQTALDVVSPLDYETTKETLRYAAQGG	480
Db	421	VTAASRDGSAVVVYYSIMSGNARGOFTLDAQTALDVVSPLDYETTKETLRYAAQGG	480
Qy	481	RPPLSNVGLVTVQVLTINDNAPIFVSTPFOATVLESVPLGLVLAQVADADAGNARL	540
Db	481	RPPLSNVGLVTVQVLTINDNAPIFVSTPFOATVLESVPLGLVLAQVADADAGNARL	540
Qy	541	EYRLAGVHADPPTINNGTGMISVAAEIDREVDYFSGVBARHGTALTAASASVTV	600
Db	541	EYRLAGVHADPPTINNGTGMISVAAEIDREVDYFSGVBARHGTALTAASASVTV	600
Qy	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRASHVITYQTISGTRRPFSTTS	660
Db	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRASHVITYQTISGTRRPFSTTS	660
Db	601	LDVNDNNPTPOPEYTVRLNEDAAVGTSVVYSAVDRASHVITYQTISGTRRPFSTTS	660
Qy	661	OSGGGLVSLALPLDYKLEROVLAVTASDGTROPTAOI.VVWVTDANTHRPVPOSSHVYN	720
Db	661	OSGGGLVSLALPLDYKLEROVLAVTASDGTROPTAOI.VVWVTDANTHRPVPOSSHVYN	720
Qy	721	VNEBRPACTTVLISATDEDTGENARITYEMEDSIPOFRIDADGAVTTQAEIDEDQVS	780
Db	721	VNEBRPACTTVLISATDEDTGENARITYEMEDSIPOFRIDADGAVTTQAEIDEDQVS	780
Qy	781	YTLAITANDNGIPKSDTTYELI.VVNDVNDNAPOFLRDSYGSYVEDVPPTSVLOISAT	840
Db	781	YTLAITANDNGIPKSDTTYELI.VVNDVNDNAPOFLRDSYGSYVEDVPPTSVLOISAT	840
Qy	841	DRDGLANGRVYFTQGGDDGDPFVBSSTGIVTLRLDEENVAQVYLRAYAVDKGMP	900
Db	841	DRDGLANGRVYFTQGGDDGDPFVBSSTGIVTLRLDEENVAQVYLRAYAVDKGMP	900
Qy	901	ARTPMEVTVTLVDVNDNPVPEODEPVDVFEENSPIGLAVARVATDPDEGTNAQIMYQI	960
Db	901	ARTPMEVTVTLVDVNDNPVPEODEPVDVFEENSPIGLAVARVATDPDEGTNAQIMYQI	960
Qy	961	VEGINIPERVOLDISGELTALVDLDYEDRPEYTVYQATSAPLYSRATVHYRLDRNDP	1020
Db	961	VEGINIPERVOLDISGELTALVDLDYEDRPEYTVYQATSAPLYSRATVHYRLDRNDP	1020
Qy	1021	PVLGNFELFNNTYNNSSSPGGAIGRVPAHDDISDLTYSPFERNELSLVTLNASTG	1080
Db	1021	PVLGNFELFNNTYNNSSSPGGAIGRVPAHDDISDLTYSPFERNELSLVTLNASTG	1080
Qy	1081	ELKLSRALDNNRPLEAIVSVSDGHSVTQAQALRYTITIDEMUTHSITLRLDMSPER	1140
Db	1081	ELKLSRALDNNRPLEAIVSVSDGHSVTQAQALRYTITIDEMUTHSITLRLDMSPER	1140
Qy	1141	FLSPLLGLFTQAVAAATLATPPDRHYVNVQCDTAPGCHILNVLSVQGPBGCGPPFL	1200
Db	1141	FLSPLLGLFTQAVAAATLATPPDRHYVNVQCDTAPGCHILNVLSVQGPBGCGPPFL	1200
Qy	1201	PSEDLQERLYNRLSLTAISQORVLPFDNCLABPCENVYRCVSLRFDSSAPFIASS	1260
Db	1201	PSEDLQERLYNRLSLTAISQORVLPFDNCLABPCENVYRCVSLRFDSSAPFIASS	1260
Qy	1261	VLFRPIHPVGGALRCRCPGFTGDYCETEVDLCYSRPGCPHRCRSREGGYTCLCRDGYTG	1320
Db	1261	VLFRPIHPVGGALRCRCPGFTGDYCETEVDLCYSRPGCPHRCRSREGGYTCLCRDGYTG	1320
Qy	1321	EHCEVSARSRCCTGVCNKGTCVNLVGGFKCPCBSGDPFKPYCOVTTBSPHASFITF	1380
Db	1321	EHCEVSARSRCCTGVCNKGTCVNLVGGFKCPCBSGDPFKPYCOVTTBSPHASFITF	1380
Qy	1381	RGLRQRFHFTLASFATKERDGLLYNGRFRNEKHDPALEVIQBOVOLTFAGSGSTTVS	1440
Db	1381	RGLRQRFHFTLASFATKERDGLLYNGRFRNEKHDPALEVIQBOVOLTFAGSGSTTVS	1440
Qy	1441	PFFVGVSIDQWHTVOLKTYNKPILLGQTGLPQGSBQKVAVTVDGCDDTVALRFGSVLG	1500
Db	1441	PFFVGVSIDQWHTVOLKTYNKPILLGQTGLPQGSBQKVAVTVDGCDDTVALRFGSVLG	1500
Qy	1501	NYSQAQGTQGGSKSLDITGPIILGGVPLPBFPRMNOFVCCMNTLQVDSHIMAD	1560
Db	1501	NYSQAQGTQGGSKSLDITGPIILGGVPLPBFPRMNOFVCCMNTLQVDSHIMAD	1560
Qy	1561	FIANGTVPGCPAKONVCDNNTCHNGTQVNMOMAFSCBPLGFGGSKCAQEMANPOHFL	1620
Db	1561	FIANGTVPGCPAKONVCDNNTCHNGTQVNMOMAFSCBPLGFGGSKCAQEMANPOHFL	1620
Qy	1621	GSSIVAHGSLPISQPMWYSLMFRTRQADGVLLQAITRGRSITTLQIREGHVLSVEGT	1680
Db	1621	GSSIVAHGSLPISQPMWYSLMFRTRQADGVLLQAITRGRSITTLQIREGHVLSVEGT	1680
Qy	1681	GLQASSLRLEPGRANDMDHAAOLALGASGPGHAILSPYGOORAGCNTGPRLHGHLIS	1740
Db	1681	GLQASSLRLEPGRANDMDHAAOLALGASGPGHAILSPYGOORAGCNTGPRLHGHLIS	1740

```

QY 1741 NITVGGIPGPAGVAGRGFGCLQGVAVSDTPEGVNSLDPDSHESINVEGCSLPPDCSN 1800
DB 1741 NITVGGIPGPAGVAGRGFGCLQGVAVSDTPEGVNSLDPDSHESINVEGCSLPPDCSN 1800
QY 1801 PCPANSYCSNDMDYSVSCDPCPYGDNCCNVCDLNPCHOSQCTKPSAPHGYTCBCPN 1860
DB 1801 PCPANSYCSNDMDYSVSCDPCPYGDNCCNVCDLNPCHOSQCTKPSAPHGYTCBCPN 1860
QY 1861 YLGPYCETRIIDQPCPRGWMGHPTGFCNCNDVSKGFPDPCNKTSGECHCKENYRPPGSP 1920
DB 1861 YLGPYCETRIIDQPCPRGWMGHPTGFCNCNDVSKGFPDPCNKTSGECHCKENYRPPGSP 1920
QY 1921 CLLCDCYPTGSLSRVCDPEDGQCPCKPGVITGRQCDRCNPPAEVTTNGCEVYDSCPRAI 1980
DB 1921 CLLCDCYPTGSLSRVCDPEDGQCPCKPGVITGRQCDRCNPPAEVTTNGCEVYDSCPRAI 1980
QY 1981 BAGIWMPPRRFGJPAAPCPKSGFGFAVHCEHGMLEPMLFNCTSTPSSLKPSAPRL 2040
DB 1981 BAGIWMPPRRFGJPAAPCPKSGFGFAVHCEHGMLEPMLFNCTSTPSSLKPSAPRL 2040
QY 2041 QNBSGLDSGRSQQLALLRNATQHTAGYFGSDVKAVALATRLAHSTQGFGLSATQ 2100
DB 2041 QNBSGLDSGRSQQLALLRNATQHTAGYFGSDVKAVALATRLAHSTQGFGLSATQ 2100
QY 2101 DVHFTENLKVSSALIDTANKHWEIIOQTEGCTAMLQHYEAVASALAQNRRHTYLSPF 2160
DB 2101 DVHFTENLKVSSALIDTANKHWEIIOQTEGCTAMLQHYEAVASALAQNRRHTYLSPF 2160
QY 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEOPDLETTYILPESVRETRPPVVRPG 2220
DB 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALRGEOPDLETTYILPESVRETRPPVVRPG 2220
QY 2221 PGEAGEPEELARORRHPELSQGEAVASYIYRTLAGLPHNYDDPKSLARVKEPIINT 2280
DB 2221 PGEAGEPEELARORRHPELSQGEAVASYIYRTLAGLPHNYDDPKSLARVKEPIINT 2280
QY 2281 PAVVISVHDDDELPRALDKPVTVQRLLETERKPCVFNHSHLVSGTGMARQGE 2340
DB 2281 PAVVISVHDDDELPRALDKPVTVQRLLETERKPCVFNHSHLVSGTGMARQGE 2340
QY 2341 VVFRSHSVSCCNHMTSPAVLMDVSRNGEILPLKLTLYVALGVTLAALLTFEPLTL 2400
DB 2341 VVFRSHSVSCCNHMTSPAVLMDVSRNGEILPLKLTLYVALGVTLAALLTFEPLTL 2400
QY 2401 IRLISNOHGIRRLTAALGALQVFLGINOADLPFACTVAILLHPLLYCTPSWALLE 2460
DB 2401 IRLISNOHGIRRLTAALGALQVFLGINOADLPFACTVAILLHPLLYCTPSWALLE 2460
QY 2461 AALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEBGNDPFCMLSTYDILI 2520
DB 2461 AALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEBGNDPFCMLSTYDILI 2520
QY 2521 WSPFAGVAVAVMSVFLYILARASCAARQGFEEKGVPVSGLOPSFAYILLSATWLANL 2580
DB 2521 WSPFAGVAVAVMSVFLYILARASCAARQGFEEKGVPVSGLOPSFAYILLSATWLANL 2580
QY 2581 LSVNSDTLLFHYLFATCNCIOGPFIFLSYVLSKEVRKALKACSRKSPDPAALTTKSTL 2640
DB 2581 LSVNSDTLLFHYLFATCNCIOGPFIFLSYVLSKEVRKALKACSRKSPDPAALTTKSTL 2640
QY 2641 TSSNCPSPYAGRLYOPYGDASGASHSTRSGKQPSYIPLLRBSLANGQGPBG 2700
DB 2641 TSSNCPSPYAGRLYOPYGDASGASHSTRSGKQPSYIPLLRBSLANGQGPBG 2700
QY 2701 DPGSLFLBEGQDQHPDSDSLSLBDDQSGSYASTHSDBEEREBEEREAAPGEG 2760
DB 2701 DPGSLFLBEGQDQHPDSDSLSLBDDQSGSYASTHSDBEEREBEEREAAPGEG 2760
QY 2761 WSLILGPAGRLPLHSTPKDGGPGKAPWPDFTTAKSSGNGAPBERLRENDALSR 2820
DB 2761 WSLILGPAGRLPLHSTPKDGGPGKAPWPDFTTAKSSGNGAPBERLRENDALSR 2820

```

```

QY 2821 EGSIGLPGSSAQPHKG 2837
DB 2821 EGSIGLPGSSAQPHKG 2837

RESULT 8
US-09-843-856-2
; Sequence 2, Application US/09843856
; Patent No. US20020034785A1
; GENERAL INFORMATION:
; APPLICANT: SOPEET, DANIEL R.
; LI, YI
; RUBEN, STEVEN M.
; TITLE OF INVENTION: CALCITONIN RECEPTOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/843,856
; FILING DATE: 30-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/970,758
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STREPER, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0660001/EKS/KMT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 568 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-843-856-2

Query Match 18.3%; Score 534; DB 9; Length 568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2388 LAALLTFEPLTLRLIRSNQHGIRRLTAALGALQVFLGINOADLPFACTVAILLH 2447
DB 33 LAALLTFEPLTLRLIRSNQHGIRRLTAALGALQVFLGINOADLPFACTVAILLH 92

QY 2448 FLVLCFSWMLLBALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEBGYN 2507
DB 93 FLVLCFSWMLLBALHYRALTEVDVNTGPMRFYTMGMVPAITGLAVGLDEBGYN 152

QY 2508 PDFCMLSYDITLINSFAGVAVAVMSVFLYILARASCAARQGFEEKGVPVSGLOPSFA 2567
DB 153 PDFCMLSYDITLINSFAGVAVAVMSVFLYILARASCAARQGFEEKGVPVSGLOPSFA 212

QY 2568 VLLLSATWMLALSVNSDTLLFHYLFATCNCIOGPFIFLSYVLSKEVRKALKACSRK 2627
DB 213 VLLLSATWMLALSVNSDTLLFHYLFATCNCIOGPFIFLSYVLSKEVRKALKACSRK 272

QY 2628 PSPDPALTTKSTLTSYNCPSPYAGRLYOPYGDASGASHSTRSGKQPSYIPLLRER 2687
DB 273 PSPDPALTTKSTLTSYNCPSPYAGRLYOPYGDASGASHSTRSGKQPSYIPLLRER 332

```

```

QY      2688 SALNPGQPPGIGADPGSLFLEGDDQHDPTDSDSLIEDQSGSYASTHSSDSEEEER 2747
DB      333 SALNPGQPPGIGADPGSLFLEGDDQHDPTDSDSLIEDQSGSYASTHSSDSEEEER 392
QY      2748 EEEEAAPFGEQGWMSILGPGARLPLHSTPKDGGPGKAPWPDGFTTAKESGNGAP 2807
DB      393 EEEBAAPFGEQGWMSILGPGARLPLHSTPKDGGPGKAPWPDGFTTAKESGNGAP 452
QY      2808 EERLRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISEKSSILRLPLEQCTGSSRG 2867
DB      453 EERLRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISEKSSILRLPLEQCTGSSRG 512
QY      2868 SASSESGRGGPPPPPPROSLQDQNGVPMIAMSITAGTVDEDSGSEFLPFNF 2921
DB      513 SSASBSGRGGPPPPPPROSLQDQNGVPMIAMSITAGTVDEDSGSEFLPFNF 566

RESULT 9
US-10-176-847-100
; Sequence 100, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-100

Query Match      16.6%; Score 485; DB 15; Length 565;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2430 INQADLPACTVIALHLHFLYCTSMALLERHLRYALTEVRDVTGPMRYMLGNGV 2489
DB      74 INQADLPACTVIALHLHFLYCTSMALLERHLRYALTEVRDVTGPMRYMLGNGV 133
QY      2490 PAFITGLAVGLDPEGYGNDPFCMLSIYDTLINSFAGPVAFVMSVFLYTLAARASCAQ 2549
DB      134 PAFITGLAVGLDPEGYGNDPFCMLSIYDTLINSFAGPVAFVMSVFLYTLAARASCAQ 193
QY      2550 RQGFKKGPVSGLQSPFAVLLLSATWLLLSVNSDTLLFHYLPATCNCIOGPFIFLSY 2609
DB      194 RQGFKKGPVSGLQSPFAVLLLSATWLLLSVNSDTLLFHYLPATCNCIOGPFIFLSY 253
QY      2610 VLSKEVKAKLKLACSRKSPDPALTTKSTLTSVNCSPYADGRLYQPYGDSASLSHST 2669
DB      254 VLSKEVKAKLKLACSRKSPDPALTTKSTLTSVNCSPYADGRLYQPYGDSASLSHST 313
QY      2670 SRSQSGQSYIFPLLRBSALNPGQPPGLGPGSLFLEGDDQHDPTDSDSLIEDD 2729
DB      314 SRSQSGQSYIFPLLRBSALNPGQPPGLGPGSLFLEGDDQHDPTDSDSLIEDD 373
QY      2730 QSGSYASTHSSDSEEEEREAAPFGEQGWMSILGPGARLPLHSTPKDGGPGKAP 2789
DB      374 QSGSYASTHSSDSEEEEREAAPFGEQGWMSILGPGARLPLHSTPKDGGPGKAP 433
QY      2790 WPGDGTAKESGNGAPERRRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISE 2849
DB      434 WPGDGTAKESGNGAPERRRNGDALSRGSLGFLPGSSAQPHKGIKKKCLPTISE 493
QY      2850 KSSLRLPLEQCTGSSRGSSASGSGRGGPPPPROSLQDQNGVPMIAMSITAGTVDE 2909
DB      494 KSSLRLPLEQCTGSSRGSSASGSGRGGPPPPROSLQDQNGVPMIAMSITAGTVDE 553

```

```

QY      2910 DSSGS 2914
DB      554 DSSGS 558

RESULT 10
US-10-264-237-2041
; Sequence 2041, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2041
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (83)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (240)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (522)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2041

Query Match      9.6%; Score 281; DB 12; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.9e-210;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2596 TCNCIOGPFIFLSYVLSKEVRKALKLACSRKSPDPALTTKSTLTSVNCSPYADGRL 2655
DB      241 TCNCIOGPFIFLSYVLSKEVRKALKLACSRKSPDPALTTKSTLTSVNCSPYADGRL 300
QY      2656 YQPYGDSAGSLHSTRSGKQPSYIFPLLRBSALNPGQPPGLGPGSLFLEGDDQHD 2715
DB      301 YQPYGDSAGSLHSTRSGKQPSYIFPLLRBSALNPGQPPGLGPGSLFLEGDDQHD 360
QY      2716 PPTSDSDSLIEDQSGSYASTHSSDSEEEEREAAPFGEQGWMSILGPGARLPLH 2775
DB      361 PPTSDSDSLIEDQSGSYASTHSSDSEEEEREAAPFGEQGWMSILGPGARLPLH 420
QY      2776 STPDGGPGKAPWPDGFTTAKESGNGAPERRRNGDALSRGSLGFLPGSSAQPH 2835
DB      421 STPDGGPGKAPWPDGFTTAKESGNGAPERRRNGDALSRGSLGFLPGSSAQPH 480
QY      2836 KGIKKKCLPTISEKSSILRLPLEQCTGSSRGSSASGSRG 2876
DB      481 KGIKKKCLPTISEKSSILRLPLEQCTGSSRGSSASGSRG 521

RESULT 11
US-09-925-300-1299
; Sequence 1299, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```

FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1299  
LENGTH: 717  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (39)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (147)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (232)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (379)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (389)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (671)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1299

Query Match  
Best Local Similarity 100.0%; Score 281; DB 10; Length 717;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2596 TCNCIGPFIPLSYVLSKVRKALKACSRKSPDPALTTKSTLTSVNCSPFYADGR 2655  
DB 390 TCNCIGPFIPLSYVLSKVRKALKACSRKSPDPALTTKSTLTSVNCSPFYADGR 449

QY 2656 YQPYGDSAGSLHSTSRSGKQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOH 2715  
DB 450 YQPYGDSAGSLHSTSRSGKQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOH 509

QY 2716 PPTDSDSLSDLDQSGSYASTHSSDSEEEEBEERAAFPQGGWMDSLGPAERLPLH 2775  
DB 510 PPTDSDSLSDLDQSGSYASTHSSDSEEEEBEERAAFPQGGWMDSLGPAERLPLH 569

QY 2776 STPDGPGPGKAPWQDPTGTAKSSGNGAPERLRENDALSRGSLGPLPGSSAOPH 2835  
DB 570 STPDGPGPGKAPWQDPTGTAKSSGNGAPERLRENDALSRGSLGPLPGSSAOPH 629

QY 2836 KGIKKKCLPTISEKSLRLPLEQCTGSSRGSSASBGSRG 2876  
DB 630 KGIKKKCLPTISEKSLRLPLEQCTGSSRGSSASBGSRG 670

RESULT 12  
US-10-017-161-1096  
Sequence 1096, Application US/10017161  
Publication No. US2003014366A1  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 084335/0152

FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/10/017,161  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: JP 2001/246789  
PRIOR FILING DATE: 2001-06-16  
NUMBER OF SEQ ID NOS: 2430  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1096  
LENGTH: 646  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-017-161-1096

Query Match  
Best Local Similarity 100.0%; Score 179; DB 12; Length 646;  
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2601 QGPFIFLSYVLSKVRKALKACSRKSPDPALTTKSTLTSVNCSPFYADGR 2660  
DB 221 QGPFIFLSYVLSKVRKALKACSRKSPDPALTTKSTLTSVNCSPFYADGR 280

QY 2661 DSAGSLHSTSRSGKQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOH 2720  
DB 281 DSAGSLHSTSRSGKQPSYIPFLAREESALNPGQPGIGDPSLFLGQDQOH 340

QY 2721 DSDLSLEDQSGSYASTHSSDSEEEEBEERAAFPQGGWMDSLGPAERLPLH 2779  
DB 341 DSDLSLEDQSGSYASTHSSDSEEEEBEERAAFPQGGWMDSLGPAERLPLH 399

RESULT 13  
US-09-764-870-331  
Sequence 331, Application US/09764870  
Patent No. US20020042386A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT214  
CURRENT APPLICATION NUMBER: US/09/764,870  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 646  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 331  
LENGTH: 219  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (119)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (123)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (145)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (189)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (192)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (196)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (211)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (214)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (219)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-870-331

Query Match 4.0%; Score 118; DB 9; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.5e-83;  
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 60

QY 627 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGLVSLALPLDYKLERQYVLA 684  
DB 61 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGLVSLALPLDYKLERQYVLA 118

RESULT 14  
US-10-125-540-331  
; Sequence 331, Application US/10125540  
; Publication No. US20030059875A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT214C1  
; CURRENT APPLICATION NUMBER: US/10/125,540  
; CURRENT FILING DATE: 2002-04-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 646  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 331  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (119)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (123)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (145)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (189)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (192)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (196)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (211)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (214)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (219)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-125-540-331

Query Match 4.0%; Score 118; DB 15; Length 219;  
Best Local Similarity 100.0%; Pred. No. 1.5e-83;

Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 60

QY 627 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGLVSLALPLDYKLERQYVLA 684  
DB 61 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGLVSLALPLDYKLERQYVLA 118

RESULT 15  
US-09-764-870-479  
; Sequence 479, Application US/09764870  
; Patent No. US20020042386A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT214  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 646  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 479  
; LENGTH: 111  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-870-479

Query Match 3.4%; Score 100; DB 9; Length 111;  
Best Local Similarity 100.0%; Pred. No. 9e-70;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 626  
DB 1 ELDRREVDYFSGVEARDHGTPALTAASASVTVLVDVNDNNPTFTQPEYTVRLNEDAVG 60

QY 627 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGL 666  
DB 61 TSVVTVSAVDRDASHVITTYQITSGNTRNRFSTSGGGGL 100

Search completed: February 11, 2004, 16:16:15  
Job time : 73 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:05:06 ; Search time 30 Seconds

(without alignments)  
4122.488 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923

Sequence: 1 MRSPTATGVPPTPPPLLL.....AGTVDSSGSEFLPFNFH 2923

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	0.7	44	1	US-07-998-003A-16
2	21	0.7	44	1	US-08-453-274B-16
3	21	0.7	44	1	US-08-453-695A-16
4	21	0.7	44	1	US-08-268-161A-16
5	21	0.7	44	2	US-08-453-702A-16
6	21	0.7	44	2	US-09-099-639-16
7	21	0.7	44	5	PCT-US93-12588-16
8	21	0.7	44	5	PCT-US95-08071-16
9	18	0.6	884	2	US-08-465-976A-2
10	18	0.6	884	2	US-08-982-412-2
11	15	0.5	43	1	US-07-998-003A-83
12	15	0.5	43	1	US-08-453-274B-83
13	15	0.5	43	1	US-08-453-695A-83
14	15	0.5	43	1	US-08-268-161A-83
15	15	0.5	43	2	US-08-453-702A-83
16	15	0.5	43	2	US-09-099-639-83
17	15	0.5	43	3	PCT-US93-12588-83
18	15	0.5	43	3	PCT-US95-08071-83
19	13	0.4	311	2	US-08-318-837-9
20	13	0.4	566	4	US-09-491-523-7
21	13	0.4	871	3	US-09-245-041-19
22	13	0.4	1211	4	US-09-491-523-5
23	13	0.4	1350	3	US-09-245-041-17
24	12	0.4	969	2	US-08-548-159-1
25	12	0.4	986	2	US-08-548-159-3
26	12	0.4	1012	4	US-08-811-481-16
27	12	0.4	1310	3	US-08-989-299-10

28	11	0.4	26	4	US-09-336-536-59	Sequence 59, Appl
29	11	0.4	44	1	US-07-998-003A-79	Sequence 79, Appl
30	11	0.4	44	1	US-08-453-274B-79	Sequence 79, Appl
31	11	0.4	44	1	US-08-453-695A-79	Sequence 79, Appl
32	11	0.4	44	1	US-08-268-161A-79	Sequence 79, Appl
33	11	0.4	44	2	US-08-453-702A-79	Sequence 79, Appl
34	11	0.4	44	3	US-09-099-639-79	Sequence 79, Appl
35	11	0.4	44	5	PCT-US93-12588-79	Sequence 79, Appl
36	11	0.4	44	5	PCT-US95-08071-79	Sequence 79, Appl
37	11	0.4	213	4	US-09-336-536-58	Sequence 58, Appl
38	11	0.4	349	4	US-09-162-524-3	Sequence 3, Appl
39	11	0.4	492	1	US-07-794-393-4	Sequence 4, Appl
40	11	0.4	492	1	US-08-001-711-4	Sequence 4, Appl
41	11	0.4	739	3	US-09-022-983-2	Sequence 2, Appl
42	11	0.4	740	3	US-09-022-983-5	Sequence 2, Appl
43	11	0.4	1001	4	US-08-884-569A-2	Sequence 2, Appl
44	11	0.4	1646	4	US-09-535-008-67	Sequence 67, Appl
45	11	0.4	1647	4	US-09-535-008-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-07-998-003A-16  
; Sequence 16, Application US/07998003A  
; Patent No. 5643781  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Marshall, O'Toole, Gerstein, Murray, &  
; STREET: 20 South Clark Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/998, 003A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5643781and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 30903  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/346-5750  
; TELEFAX: 312/984-9740  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-998-003A-16

Query Match 0.7% Score 21; DB 1; Length 44;

Best Local Similarity 100.0%; Pred. No. 5.4e-11; Mismatches 0; Gaps 0;

Matches 21; Conservative 0; Indels 0; Gaps 0;  
Cy 483 PLSNVSGLVTVQVADINDNAP 503  
Db 24 PLSNVSGLVTVQVADINDNAP 44



RESULT 2  
US-08-453-274B-16  
; Sequence 16, Application US/08453274B  
; Patent No. 5663300  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,274B  
; FILING DATE: 30-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5663300and, Greeta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 32660  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-453-274B-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
Db 24 PLSNVSGLVTVQVLDINDNAP 44

RESULT 3  
US-08-453-695A-16  
; Sequence 16, Application US/08453695A  
; Patent No. 5708143  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESSER: Borun  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,695A

FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greeta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-695A-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
Db 24 PLSNVSGLVTVQVLDINDNAP 44

RESULT 4  
US-08-268-161A-16  
; Sequence 16, Application US/08268161A  
; Patent No. 5798224  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESSER: Borun  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/268,161A  
; FILING DATE: June 27, 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Young J. Suh  
; REGISTRATION NUMBER: P-41,337  
; REFERENCE/DOCKET NUMBER: 27866/32149  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-268-161A-16

Query Match 0.7%; Score 21; DB 1; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 5

US-08-453-702A-16  
Sequence 16, Application US/08453702A  
Patent No. 5891706  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-702A-16

Query Match 0.7%; Score 21; DB 2; Length 44;

Best Local Similarity 100.0%; Pred. No. 5,4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 6

US-09-099-639-16  
Sequence 16, Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161  
FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 6262237and  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-099-639-16

Query Match 0.7%; Score 21; DB 3; Length 44;

Best Local Similarity 100.0%; Pred. No. 5,4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 7

PCT-US93-12588-16  
Sequence 16, Application PCT/US9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-12588-16

Query Match 0.7%; Score 21; DB 5; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 8

PCT-US95-08071-16  
Sequence 16, Application PC/TUS9508071  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08071  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE: 23 DEC 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-08071-16

Query Match 0.7%; Score 21; DB 5; Length 44;  
Best Local Similarity 100.0%; Pred. No. 5.4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 PLSNVSGLVTVQVLDINDNAP 503  
DB 24 PLSNVSGLVTVQVLDINDNAP 44

## RESULT 9

US-08-465-976A-2  
Sequence 2, Application US/08465976A  
Patent No. 5858632  
GENERAL INFORMATION:  
APPLICANT: SOPPET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: RUBEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN GILFILLAN, CECCHI  
ADDRESSEE: STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NJ  
COUNTRY: US  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,976A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY F  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-444  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 994-1700  
TELEFAX: (201) 994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 884 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-976A-2

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGLDPFGYGNPDCW 2512  
DB 220 GLAVGLDPFGYGNPDCW 237

## RESULT 10

US-08-982-412-2  
Sequence 2, Application US/08982412  
Patent No. 5958729  
GENERAL INFORMATION:  
APPLICANT: SOPPET, DANIEL R  
APPLICANT: LI, YI  
APPLICANT: ROSEN, CRAIG A  
APPLICANT: RUBEN, STEVEN M  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/982,412  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOKES, ANDERS A  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF181PCT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 884 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-982-412-2

Query Match 0.6%; Score 18; DB 2; Length 884;  
Best Local Similarity 100.0%; Pred. No. 3,9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGLDPEGYGNPDFCW 2512  
Db 220 GLAVGLDPEGYGNPDFCW 237

RESULT 11  
US-07-998-003A-83  
Sequence 83, Application US/07998003A  
Patent No. 5643781  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 20 South Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/998,003A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5643781and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 30903  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/346-5750  
TELEFAX: 312/984-9740  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-998-003A-83

Query Match 0.5%; Score 15; DB 1; Length 43;

Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 791 GIPOKSDTYLTLYLV 805  
Db 21 GIPOKSDTYLTLYLV 35

RESULT 12  
US-08-453-274B-83  
Sequence 83, Application US/08453274B  
Patent No. 5663300  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-274B-83

Query Match 0.5%; Score 15; DB 1; Length 43;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GIPOKSDTYLTLYLV 805  
Db 21 GIPOKSDTYLTLYLV 35

RESULT 13  
US-08-453-695A-83  
Sequence 83, Application US/08453695A  
Patent No. 5708143  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-83

Query Match      0.5%; Score 15; DB 1; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      791 GIPKSDTYLEILV 805
DB      21 GIPKSDTYLEILV 35

RESULT 14
US-08-268-161A-83
; Sequence 83, Application US/08268161A
; Patent No. 5798224
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/268,161A
; FILING DATE: June 27, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-268-161A-83

Query Match      0.5%; Score 15; DB 1; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      791 GIPKSDTYLEILV 805
DB      21 GIPKSDTYLEILV 35

RESULT 15
US-08-453-702A-83
; Sequence 83, Application US/08453702A
; Patent No. 5891706
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,702A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5891706and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32657
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-702A-83

Query Match      0.5%; Score 15; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      791 GIPKSDTYLEILV 805
DB      21 GIPKSDTYLEILV 35
```

Search completed: February 11, 2004, 16:09:43  
Job time : 31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:03:46 ; Search time 39 Seconds

(without alignments)  
7207.719 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923

Sequence: 1 MRSFATGVLPFTPPPLLL.....AGTVDEDSGSEFLFNFLL 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	0.8	3034	2 T14119	seven-pass transme
2	18	0.6	1364	2 T00250	MEGF2 protein - hu
3	12	0.4	491	2 JC6197	stromelysin 3 (BC
4	12	0.4	608	2 T06632	hypothetical prote
5	12	0.4	1015	2 UC5263	transmembrane tyro
6	12	0.4	1015	2 UC5062	phogrin precursor
7	12	0.4	1032	2 T18293	guanylate kinase-i
8	12	0.4	1309	1 S35484	peptidyl-dipeptida
9	11	0.4	75	2 T06013	hypothetical prote
10	11	0.4	106	2 T06479	proline/leucine-ri
11	11	0.4	182	2 A40959	secretin precursor
12	11	0.4	134	2 UC7236	receptor-activit
13	11	0.4	189	2 JC7262	neuromodulin - gol
14	11	0.4	213	2 J00075	hypothetical prote
15	11	0.4	316	2 G96513	transcription init
16	11	0.4	393	2 T43401	NNE-4AG - human (f
17	11	0.4	412	2 G02453	hypothetical prote
18	11	0.4	445	2 T47813	stromelysin 3 (BC
19	11	0.4	492	2 A44399	80K protein H prec
20	11	0.4	527	2 A32469	prostaglandin G/H
21	11	0.4	602	2 S69198	cyclooxygenase 1 -
22	11	0.4	617	2 S39782	actin-binding prot
23	11	0.4	617	2 S42719	WYD protein inhib
24	11	0.4	675	2 T03744	histidine rich cal
25	11	0.4	699	2 A54660	Fas-binding protei
26	11	0.4	736	2 T03849	Fas-binding protei
27	11	0.4	740	2 T03847	protein kinase (EC
28	11	0.4	740	2 H54024	protein kinase (EC
29	11	0.4	777	2 B54024	

30	11	0.4	777	2 F54024	protein kinase (EC
31	11	0.4	779	2 B54024	protein kinase (EC
32	11	0.4	783	2 A55817	cyclin-dependent k
33	11	0.4	1001	2 S30385	G9a protein - huma
34	11	0.4	1022	2 T53078	homeotic gene regu
35	11	0.4	1078	2 T42712	myelin transcripti
36	11	0.4	1184	2 A55184	fibulin-2 precursor
37	11	0.4	1467	2 T48162	hypothetical prote
38	11	0.4	1613	2 S39059	protein BRG1 - hum
39	11	0.4	1647	2 S45252	SNP2beta protein -
40	11	0.4	1753	2 T00350	hypothetical prote
41	11	0.4	2009	2 S49764	SEC7 protein - yea
42	10	0.3	81	2 T48398	hypothetical prote
43	10	0.3	81	2 T47289	hypothetical prote
44	10	0.3	105	2 A27471	homeotic protein R
45	10	0.3	119	2 A03314	homeotic protein m

## ALIGNMENTS

RESULT 1  
T14119  
seven-pass transmembrane receptor protein precursor - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C/Accession: T14119  
R:Hadjantonakis, A.K.; Formstone, C.J.; Little, P.F.R.  
A/Description: The Celst family of novel evolutionarily conserved seven-pass transmembran  
A/Reference number: Z17881  
A/Accession: T14119  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-3034 <HAD>  
A/Cross-references: EMBL:AF031572; NID:G3800735; PID:G3800736; PIDN:AAC68836.1  
C/Genetics:  
A/Gene: Celst1  
A/Map position: 15  
C/Keywords: transmembrane protein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-3034/Product: seven-pass transmembrane receptor protein #status predicted <MNT>

Query Match 0.8%; Score 23; DB 2; Length 3034;  
Best Local Similarity 100.0%; Pred. No. 9.3e-12;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1219 ISARVLPFDNICTLRPCENYM 1241  
Db 1309 ISARVLPFDNICTLRPCENYM 1331  
|||||

RESULT 2  
T00250  
MEGF2 protein - human (fragment)  
C/Species: Homo sapiens (hmn)  
C/Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
C/Accession: T00250  
R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.  
A/Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs  
A/Reference number: Z14126; MUID:98360089; PMID:9693030  
A/Accession: T00250  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1364 <YAK>  
A/Cross-references: EMBL:AB011536; NID:G3449297; PIDN:BAA32464.1; PID:G3449298  
A/Experimental source: brain; Clone HG1044  
C/Genetics:  
A/Gene: MEGF2  
A/Map position: 3p21.2-p24.1  
C/Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homo  
F:1-28/Domain: EGF homology (fragment) <EGF>

F:32-66/Domain: EGF homology <EGF1>  
F:124-169/Domain: laminin-type EGF-like homology <LEG>

Query Match 0.4%; Score 18; DB 2; Length 1364;  
Best Local Similarity 100.0%; Pred. No. 2e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2495 GLAVGDPGKGNPFCM 2512  
DB 710 GLAVGDPGKGNPFCM 727

#### RESULT 3

JC6197  
stromelysin 3 (EC 3.4.24.-) - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 17-Mar-1999  
C/Accession: J06197  
R:Okada, A.; Saez, S.; Mismul, Y.; Basset, P.  
Gene 185, 187-193, 1997  
A/Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin an  
A/Reference number: J06197; MUID:97208872; PMID:9055814  
A/Contents: Skin wounds  
A/Accession: J06197  
A/Molecule type: mRNA  
A/Residues: 1-491 <OKA>  
A/Cross-references: GB:U46034  
C/Comment: This protein is a member of the matrix metalloproteinase family.  
C/Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloprotei  
C/Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F:52-261/Domain: matrix metalloproteinase homology <MMP>  
F:291-483/Domain: hemopexin repeat homology <HPX>  
F:84,218,222,228/Binding site: zinc, catalytic (His)  
F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted  
F:219/Active site: Glu #status predicted

Query Match 0.4%; Score 12; DB 2; Length 491;  
Best Local Similarity 100.0%; Pred. No. 0.03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PULLLLLLLLP 27  
DB 20 PULLLLLLLLP 31

#### RESULT 4

T06632  
hypothetical protein T20K18.90 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
C/Accession: T06632  
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkee, W.; Stiekema, W.; Bancroft, I.; Mew  
submitted to the Protein Sequence Database, April 1999  
A/Reference number: Z15190  
A/Accession: T06632  
A/Molecule type: DNA  
A/Residues: 1-608 <BEV>  
A/Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.90  
A/Experimental source: cultivar Columbia; BAC clone T20K18  
C/Genetic:  
A/Gene: ATSP:T20K18.90  
A/Map position: 4  
A/Intons: 202/3; 216/2; 221/3; 241/3; 292/2; 471/1

Query Match 0.4%; Score 12; DB 2; Length 608;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2742 SEEBEBEBEBA 2753  
DB 97 SEEBEBEBEBA 108

RESULT 5  
JCS263  
transmembrane tyrosine phosphatase-like protein, ICAR - human

C/Species: Homo sapiens (man)  
C/Date: 25-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 20-Jun-2000  
C/Accession: JCS263  
R:Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.  
Biochem. Biophys. Res. Commun. 229, 402-411, 1996  
A/Title: ICAR, a novel member of a new family of transmembrane, tyrosine phosphatase-11  
A/Reference number: JCS263; MUID:97127415; PMID:8954911  
A/Accession: JCS263  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-1015 <SMI>  
A/Cross-references: GB:Y08569; NID:G1644377; PIDN:CA69880.1; PID:G1644378  
C/Comment: This protein has an intracellular protein tyrosine phosphatase like protein.  
C/Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata  
C/Keywords: phosphoprotein  
F:770-994/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:945/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:951/Binding site: substrate phosphate (Arg) #status predicted

Query Match 0.4%; Score 12; DB 2; Length 1015;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PULLLLLLLLP 27  
DB 6 PULLLLLLLLP 17

#### RESULT 6

JCS062  
phogrin precursor - human  
N/Contains: protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type  
C/Species: Homo sapiens (man)  
C/Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 21-Jun-2002  
C/Accession: JCS062; JCS263; T46903  
R:Kawasaki, E.; Hutton, J.C.; Eisenbach, G.S.  
Biochem. Biophys. Res. Commun. 227, 440-447, 1996  
A/Title: Molecular cloning and characterization of the human transmembrane protein tyros  
A/Reference number: JCS062; MUID:97032784; PMID:8878534  
A/Contents: Islet  
A/Accession: JCS062  
A/Molecule type: mRNA  
A/Residues: 1-1015 <KAW>  
A/Cross-references: GB:U6702; NID:G1620663; PIDN:AA050742.1; PID:G1620664  
R:Smith, P.D.; Barker, K.T.; Wang, J.; Lu, Y.J.; Shipley, J.; Crompton, M.R.  
Biochem. Biophys. Res. Commun. 229, 402-411, 1996  
A/Title: ICAR, a novel member of a new family of transmembrane, tyrosine phosphatase-11  
A/Reference number: JCS263; MUID:97127415; PMID:8954911  
A/Accession: JCS263  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-207, 'S', 209-246, 'G', 248-322, 'R', 324, 'N', 326-1015 <SMI>  
A/Cross-references: GB:Y08569; NID:G1644377; PIDN:CA69880.1; PID:G1644378  
R:Amoroge, W.; Wilkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, February 2000  
A/Reference number: Z24134  
A/Accession: T46903  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 'DA', 714, 771-1015 <AAA>  
A/Cross-references: EMBL:AL157451  
C/Comment: This protein has an intracellular protein tyrosine phosphatase like protein.  
C/Genetic:  
A/Note: DKP2G761A0712.1  
C/Superfamily: protein-tyrosine-phosphatase, receptor type N; protein-tyrosine-phosphata  
C/Keywords: phosphoprotein; phosphoric monoester hydrolyase; transmembrane protein; tyros  
F:1-1015/Domain: signal sequence #status predicted <SIG>  
F:18-10/Product: phogrin #status predicted <MAT>  
F:615-639/Domain: transmembrane #status predicted <TM>



F:770-994/Domain: protein-tyrosine-phosphatase homology <PTP2>  
 F:945/Active site: Cys (phosphotyrosine intermediate) #status predicted  
 F:951/Binding site: substrate phosphate (Arg) #status predicted

Query Match 0.4%; Score 12; DB 2; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 PLLLLLLLLLPP 27  
 |||||  
 DB 6 PLLLLLLLLLPP 17

RESULT 7  
 T18293  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18293  
 R:Yao, L.; Hata, Y.; Takai, Y.  
 Submitted to the EMBL Data Library, October 1998  
 A:Description: Membrane-associated guanylate kinase-interacting protein.  
 A:Reference number: Z18862  
 A:Accession: T18293  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1032 <YAO>  
 A:Cross-references: EMBL:AF102853; NID:G4151804; PID:G4151805; PIDN:RAD04567.1

Query Match 0.4%; Score 12; DB 2; Length 1032;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2743 EEEEEEEEEEA 2754  
 |||||  
 DB 877 EEEEEEEEEEA 888

RESULT 8  
 S35484  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S35484; A23455; A18700; A38655; A49726; S17509  
 R:Thekkumkara, T.J.; Livingston III, W.; Kumar, R.S.; Sen, G.C.  
 Nucleic Acids Res. 20, 683-687, 1992  
 A:Title: Use of alternative polyadenylation sites for tissue-specific transcription of t  
 A:Reference number: S35484; MUID:92178960; PMID:1311831  
 A:Accession: S35484  
 A:Molecule type: mRNA  
 A:Residues: 1-1309 <THB>  
 A:Cross-references: EMBL:X62551  
 R:Wata, K.; Blacher, R.; Softer, R.L.; Lai, C.Y.  
 Arch. Biochem. Biophys. 227, 188-201, 1983  
 A:Reference number: A23455; MUID:84051289; PMID:6314908  
 A:Accession: A23455  
 A:Molecule type: protein  
 A:Residues: 34-47, 'N', 49-55 <IWA>  
 A:Experimental source: lung  
 R:Wata, K.; Lai, C.Y.; El-Dorty, H.A.; Softer, R.L.  
 Biochem. Biophys. Res. Commun. 107, 1097-1103, 1982  
 A:Title: The NH2- and COOH-terminal sequences of the angiotensin-converting enzyme isozym  
 A:Reference number: A50107; MUID:83048249; PMID:6291514  
 A:Accession: A18700  
 A:Molecule type: protein  
 A:Residues: 34-44, 754-755, 'V', 757 <IWA>  
 R:Kumar, R.S.; Thekkumkara, T.J.; Sen, G.C.  
 J. Biol. Chem. 266, 3854-3862, 1991  
 A:Title: The mRNAs encoding the two angiotensin-converting isozymes are transcribed from  
 A:Reference number: A38655; MUID:91139683; PMID:1847388  
 A:Accession: A38655  
 A:Molecule type: DNA

A:Residues: 1-88 <KDM>  
 A:Cross-references: GB:M5879  
 R:Ranchandran, R.; Sen, G.C.; Misano, K.; Sen, I.  
 J. Biol. Chem. 269, 2125-2130, 1994  
 A:Title: Regulated cleavage-secretion of the membrane-bound angiotensin-converting enzyme  
 A:Reference number: A49726; MUID:94124566; PMID:8294466  
 A:Accession: A49726  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1236-1258 <RAM>  
 A:Experimental source: testis  
 R:Kitley, T.L.  
 Biochem. J. 278, 375-380, 1991  
 A:Title: The Mg(2+)-ATPase of rabbit skeletal-muscle transverse tubule is a highly glycos  
 A:Reference number: S17509; MUID:91378880; PMID:1654880  
 A:Accession: S17509  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 34-55 <KIR>  
 A:Comment: This enzyme converts angiotensin I to angiotensin II in presence of divalent c  
 ver, the enzyme has been found also in renal tubules and intestinal mucosa.  
 C:Superfamily: mammalian peptidyl-dipeptidase A  
 C:Keywords: alternative splicing; blood pressure control; chloride; glycoprotein; interest  
 F:1-33/Domain: signal sequence #status predicted <SIG>  
 F:34-1309/Product: peptidyl-dipeptidase A, pulmonary #status experimental <MAT>  
 F:159,79,150,322,448,512,680,698,717,945,1194/Binding site: carbohydrate (asn) (covalent)

Query Match 0.4%; Score 12; DB 1; Length 1309;  
 Best Local Similarity 100.0%; Pred. No. 0.07;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 PPPLLLLLLLL 24  
 |||||  
 DB 15 PPPLLLLLLLL 26

RESULT 9  
 T06013  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06013  
 R:Beran, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z15382  
 A:Accession: T06013  
 A:Molecule type: DNA  
 A:Residues: 1-75 <BRV>  
 A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.100  
 A:Experimental source: cultivar Columbia; BAC clone T25K17  
 C:Genetics:  
 A:Gene: ATSP:T25K17.100  
 A:Map position: 4

Query Match 0.4%; Score 11; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 0.052;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2743 EEEEEEEER 2753  
 |||||  
 DB 13 EEEEEEEER 23

RESULT 10  
 T06479  
 C:Species: Pisum sativum (garden pea)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Nov-1999  
 C:Accession: T06479  
 R:Rodriguez-Concepcion, M.; Perez-Garcia, A.; Beltrán, J.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: Isolation of cDNAs which accumulate during pea (Pisum sativum L.) early fr  
 A:Reference number: Z15708

A:Accession: T06479  
A:Species: Rattus norvegicus (Norway rat)  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-106 <R0D>  
A:Cross-references: EMBL:267873; NID:g1213129; PIDN:CAA1780.1; PID:e208986  
F:1-1//Domain: signal sequence #status predicted <Sig>  
F:18-106/Product: proline/leucine-rich protein #status predicted <MAT>

Query Match  
Best Local Similarity 100.0%; Score 11; DB 2; Length 106;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LLLLLPPPL 30  
DB 73 LLLLLPPPL 83

RESULT 11  
A:Accession: A40959  
A:Species: Rattus norvegicus (Norway rat)  
C:Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: A40886; A40959; A35094; A32544  
R:Itou, N.; Furuya, T.; Ozaki, K.; Onita, W.; Kawasaki, T.  
J. Biol. Chem. 266, 12595-12598, 1991  
A:Title: The secretin precursor gene. Structure of the coding region and expression in t  
A:Reference number: A40886; MUID:91286291; PMID:2061329  
A:Accession: A40886  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-134 <RTO>  
A:Cross-references: GB:M63984; NID:g206889; PIDN:AAA42127.1; PID:g206890  
R:Kopin, A.S.; Wheeler, M.B.; Nishitani, J.; McBride, E.W.; Chang, T.; Chey, W.Y.; Lett  
Proc. Natl. Acad. Sci. U.S.A. 88, 5335-5339, 1991  
A:Title: The secretin gene: evolutionary history, alternative splicing, and development  
A:Reference number: A40959; MUID:91271384; PMID:1711228  
A:Accession: A40959  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-134 <R0P>  
A:Cross-references: GB:M64033; NID:g206891; PIDN:AAA42128.1; PID:g206892  
R:Kopin, A.S.; Wheeler, M.B.; Lettler, A.B.  
Proc. Natl. Acad. Sci. U.S.A. 87, 2289-2303, 1990  
A:Title: Secretin: structure of the precursor and tissue distribution of the mRNA.  
A:Reference number: A35094; MUID:90192795; PMID:2315322  
A:Accession: A35094  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-134 <R0P2>  
A:Cross-references: GB:M31495; NID:g206887; PIDN:AAA42126.1; PID:g206888  
R:Gosden, D.; Vandermere, A.; Vandermere-Piret, M.C.; Rathe, J.; Cauvin, A.; Robberech  
Biochem. Biophys. Res. Commun. 160, 862-867, 1989  
A:Title: Isolation and primary structure of rat secretin.  
A:Reference number: A32544; MUID:89246545; PMID:2719704  
A:Accession: A32544  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 33-59 <GOS>  
C:Superfamily: glucagon  
C:Keyword: duplication

Query Match  
Best Local Similarity 100.0%; Score 11; DB 2; Length 134;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25  
DB 8 PLLLLLLLLL 18

RESULT 12  
JC7236  
receptor-activity-modifying protein 2 - rat

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: JC7236  
R:Nagase, T.; Mkyama, M.; Sugawara, A.; Mori, K.; Yahata, K.; Kasahara, M.; Suganami, T  
Biochem. Biophys. Res. Commun. 270, 89-93, 2000  
A:Title: Rat receptor-activity-modifying proteins (RAMPs) for adrenomedullin/CGRP recept  
A:Reference number: JC7235  
A:Accession: JC7236  
A:Molecule type: mRNA  
A:Residues: 1-182 <NAG>  
A:Cross-references: DDBJ:AB030943  
C:Comment: This protein is a transmembrane glycoprotein having roles in renal pathophys  
as the adrenomedullin receptor.  
C:Keywords: glycoprotein; transmembrane protein

Query Match  
Best Local Similarity 100.0%; Score 11; DB 2; Length 182;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25  
DB 29 PLLLLLLLLL 39

RESULT 13  
JC7262  
receptor activity modifying protein 2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: JC7262  
R:Ono, Y.; Okano, I.; Kojima, M.; Okada, K.; Kangawa, K.  
Biochem. Biophys. Res. Commun. 271, 197-202, 2000  
A:Title: Decreased gene expression of adrenomedullin receptor in mouse lungs during seps  
A:Reference number: JC7261  
A:Accession: JC7262  
A:Molecule type: mRNA  
A:Residues: 1-189 <ONO>  
A:Cross-references: GB:AF209906  
C:Comment: This protein interacts with calcitonin receptor-like receptor, modifies its lig  
lation.  
C:Genetics:  
A:Gene: Yamp2  
C:Keywords: transmembrane protein

Query Match  
Best Local Similarity 100.0%; Score 11; DB 2; Length 189;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PLLLLLLLLL 25  
DB 29 PLLLLLLLLL 39

RESULT 14  
UQ0075  
neuromodulin - goldfish  
N:Alternate names: B-50; F1; GAP-43; PP46  
C:Species: Carassius auratus (goldfish)  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 21-Jul-2000  
C:Accession: UQ0075  
R:Labate, M.E.; Skene, J.H.P.  
Neuron 3, 299-310, 1989  
A:Title: Selective conservation of GAP-43 structure in vertebrate evolution.  
A:Reference number: UQ0075; MUID:90380372; PMID:2641999  
A:Accession: UQ0075  
A:Molecule type: mRNA  
A:Residues: 1-213 <LAB>  
A:Cross-references: GB:M26250; NID:g212953; PIDN:AAA03010.1; PID:g212954  
A:Note: This protein contains potential sites for fatty acylation and membrane attachment  
C:Comment: GAP-43 is a major growth cone membrane protein whose expression is widely cor  
C:Comment: Features of the protein are most highly conserved in vertebrate evolution.  
C:Comment: The amino end (1-57) has a strictly conserved domain including protein-protein  
C:Superfamily: histone H1

C:Keywords: calmodulin binding; lipoprotein; phosphoprotein; thiolester bond  
 F:3,4/Binding site: palmitate (Cys) (covalent) #status predicted  
 F:81/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 0.4%; Score 11; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 EEEEEEEEA 2753

DB 161 EEEEEEEEA 171

# RESULT 15

G96513  
 hypothetical protein T3P24.8 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: G96513

R:Theologis, A.; Ecker, J.R.; Palm, C.U.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96513

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:AE005173; NID:g9993350; PIDN:AG11423.1; GSPDB:GN00141

C:Genetics:

A:Gene: T3P24.8

A:Map position: 1

Query Match 0.4%; Score 11; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 0.18;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2742 SEEEEEEER 2752

DB 220 SEEEEEEER 230

Search completed: February 11, 2004, 16:09:00  
 Job time : 40 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:58:30 ; Search time 25 Seconds

(without alignments)  
5498.362 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923  
Sequence: 1 MRSPTATGVLPTPPPLLL.....AGTVDEDSGSRFLPFNLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	2923	1	CLR2_HUMAN
2	187	6.4	2144	1	CLR2_RAT
3	168	5.7	2920	1	CLR1_MOUSE
4	23	0.8	3034	1	CLR1_MOUSE
5	20	0.7	3014	1	CLR1_HUMAN
6	18	0.6	3301	1	CLR3_MOUSE
7	18	0.6	3312	1	CLR3_HUMAN
8	18	0.6	3313	1	CLR3_RAT
9	15	0.5	3298	1	PC16_HUMAN
10	13	0.4	1211	1	ATG2_HUMAN
11	13	0.4	1211	1	MYC1_CYPCA
12	13	0.4	1967	1	CD87_DROME
13	13	0.4	1967	1	RLA0_METRA
14	12	0.4	1013	1	PTPX_MACNE
15	12	0.4	1015	1	PTPX_HUMAN
16	12	0.4	1121	1	MYT1_HUMAN
17	12	0.4	1310	1	ACE_RABIT
18	12	0.4	1377	1	CBPD_MOUSE
19	11	0.4	134	1	SECR_RAT
20	11	0.4	189	1	RMP2_MOUSE
21	11	0.4	213	1	NEOM_CARAU
22	11	0.4	234	1	MDC2_HUMAN
23	11	0.4	393	1	T2D6_SCHPO
24	11	0.4	398	1	SHBG_RABIT
25	11	0.4	399	1	SHBG_PHOSU
26	11	0.4	492	1	MM11_MOUSE
27	11	0.4	527	1	GL1P_HUMAN
28	11	0.4	559	1	GL6S_CAPII
29	11	0.4	574	1	SEN3_HUMAN
30	11	0.4	597	1	SIL1_PANTR
31	11	0.4	602	1	PGH1_RAT
32	11	0.4	617	1	ABP1_SACEX
33	11	0.4	699	1	SRCH_HUMAN

34	11	0.4	731	1	DAXX_RAT	Q9yib2	rattus norv
35	11	0.4	736	1	DAXX_CERAB	Q18805	cercopithe
36	11	0.4	739	1	DAXX_MOUSE	Q35613	mus musculu
37	11	0.4	740	1	DAXX_HUMAN	Q9uer7	homo sapien
38	11	0.4	741	1	PLO3_MOUSE	Q9r0e1	mus musculu
39	11	0.4	832	1	SM4B_HUMAN	Q9npr2	homo sapien
40	11	0.4	851	1	STR8_MOUSE	Q8k031	mus musculu
41	11	0.4	979	1	CLS1_MOUSE	Q9ep12	mus musculu
42	11	0.4	981	1	CLS1_HUMAN	Q94985	homo sapien
43	11	0.4	1001	1	PTPX_HUMAN	P80560	mus musculu
44	11	0.4	1004	1	PTPX_RAT	Q63475	rattus norv
45	11	0.4	1007	1	CHC2_HUMAN	Q9y514	homo sapien

## ALIGNMENTS

RESULT 1  
CLR2\_HUMAN STANDARD; PRT; 2923 AA.  
ID CLR2\_HUMAN  
AC Q9HCU4; Q92566;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Cadherin-1-like protein with epidermal growth factor-1-like domains  
DE growth factor-1-like 2 (Multiple epidermal growth factor-1-like domains  
DE 3) (Fleming 1).  
GN CELSR2 OR CDH10 OR EGF2 OR MEGF3 OR KIAA0279.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363102; PubMed=10907856;  
RA Vincent U.B., Skang U., Scherer S.W.;  
RT "The human homologue of flamingo, EGF2, encodes a brain-expressed  
RT large cadherin-like protein with epidermal growth factor-1-like domains,  
RT and maps to chromosome 1p13.3-p21.1.";  
RL DNA Res. 7:233-235(2000).  
RN [2]  
RP SEQUENCE OF 516-2923 FROM N.A.  
RX TISSUE=Brain;  
RC MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y.,  
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
CC - FUNCTION: Receptor that may have an important role in cell/cell  
CC signaling during nervous system formation.  
CC - SUBCELLULAR LOCATION: Integral membrane protein.  
CC - TISSUE SPECIFICITY: Highest expression in brain and testis.  
CC - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC - SIMILARITY: Contains 9 cadherin domains.  
CC - SIMILARITY: Contains 8 EGF-like domains.  
CC - SIMILARITY: Contains 2 laminin G-like domains.  
CC - SIMILARITY: Contains 1 GPS domain.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage is for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AF234887; AAC00080.1; -  
CC EMBL; D87469; BA13407.1; -  
CC HSSP; P15116; INCI.  
CC DR GENSUP; KIAA0279; KIAA0279.  
CC DR MIM; 604265; -  
CC

DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0004930; F:G-protein coupled receptor activity; NAS.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_CA.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; laminin\_BGF.  
 DR InterPro; IPR001791; laminin\_G.  
 DR InterPro; IPR002023; PD\_cys-rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GFS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00053; laminin\_EGF; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00303; GFS; 1.  
 DR SMART; SM00282; Hormr; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 7.  
 DR PROSITE; PS0266; CADHERIN\_2; 9.  
 DR PROSITE; PS00221; GFS; 1.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00649; G PROTEIN RECP\_F2\_1; FALSE NEG.  
 DR PROSITE; PS00650; G PROTEIN RECP\_F2\_2; FALSE NEG.  
 DR PROSITE; PS00227; G PROTEIN RECP\_F3\_1.  
 DR PROSITE; PS00261; G PROTEIN RECP\_F4\_1.  
 DR PROSITE; PS00261; LAM G DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ TYPE\_EGF; 1.  
 KM G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KM EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 KM Developmental protein; Hydroxylation; Signal.  
 FT CHAIN 1 2923  
 FT SIGNAL 31  
 FT DOMAIN 32 2380  
 FT TRANSSEM 2381 2401  
 FT TRANSSEM 2402 2416  
 FT TRANSSEM 2417 2437  
 FT TRANSSEM 2438 2438  
 FT TRANSSEM 2439 2459  
 FT TRANSSEM 2460 2480  
 FT TRANSSEM 2481 2501  
 FT TRANSSEM 2502 2519  
 FT TRANSSEM 2520 2540  
 FT TRANSSEM 2541 2560  
 FT TRANSSEM 2561 2581  
 FT TRANSSEM 2582 2591  
 FT TRANSSEM 2592 2612  
 FT TRANSSEM 2613 2923  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 505  
 FT DOMAIN 506 610  
 FT DOMAIN 611 712  
 FT DOMAIN 713 815  
 FT DOMAIN 816 921  
 FT DOMAIN 922 1023  
 FT DOMAIN 1028 1146  
 FT DOMAIN 1228 1286  
 FT DOMAIN 1288 1324  
 FT DOMAIN 1328 1366  
 FT DOMAIN 1367 1571

FT DOMAIN 1574 1610  
 FT DOMAIN 1614 1791  
 FT DOMAIN 1793 1828  
 FT DOMAIN 1829 1867  
 FT DOMAIN 1883 1922  
 FT DOMAIN 1923 1955  
 FT DOMAIN 2316 2368  
 FT DOMAIN 2743 2752  
 FT DISULFID 1232 1243  
 FT DISULFID 1237 1274  
 FT DISULFID 1276 1285  
 FT DISULFID 1292 1303  
 FT DISULFID 1297 1312  
 FT DISULFID 1314 1323  
 FT DISULFID 1332 1343  
 FT DISULFID 1337 1353  
 FT DISULFID 1355 1365  
 FT DISULFID 1378 1389  
 FT DISULFID 1578 1589  
 FT DISULFID 1583 1598  
 FT DISULFID 1600 1609  
 FT DISULFID 1797 1808  
 FT DISULFID 1802 1817  
 FT DISULFID 1819 1828  
 FT DISULFID 1832 1843  
 FT DISULFID 1837 1855  
 FT DISULFID 1857 1866  
 FT DISULFID 1887 1899  
 FT DISULFID 1889 1906  
 FT DISULFID 1908 1921  
 FT DISULFID 1924 1936  
 FT DISULFID 1926 1943  
 FT DISULFID 1945 1954  
 FT MOD\_RES 1591 1591  
 FT MOD\_RES 1810 1810  
 FT CARBOHYD 486 486  
 FT CARBOHYD 557 557  
 FT CARBOHYD 701 701  
 FT CARBOHYD 1036 1036  
 FT CARBOHYD 1076 1076  
 FT CARBOHYD 1182 1182  
 FT CARBOHYD 1212 1212  
 FT CARBOHYD 1501 1501  
 FT CARBOHYD 1565 1565  
 FT CARBOHYD 1741 1741  
 FT CARBOHYD 1827 1827  
 FT CARBOHYD 1900 1900  
 FT CARBOHYD 2024 2024  
 FT CARBOHYD 2061 2061  
 FT CARBOHYD 2323 2323  
 FT CARBOHYD 2345 2345  
 SQ SEQUENCE 2923 AA; 317447 MW; 3827579315158BD8 CRC64;

Query Match 100.0%; Score 2923; DB 1; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPATGVLPPEPPLILLLLPPPLIGQVPCSSISGRGSSGACAPMGWLCPS 60  
 DB 1 MSPATGVLPPEPPLILLLLPPPLIGQVPCSSISGRGSSGACAPMGWLCPS 60  
 QY 61 SASNLWLYTSRCRDAGTELTGHLVPHDGLRWVCESSEAHIPLPAPAECCPMSCRLLGIG 120  
 DB 61 SASNLWLYTSRCRDAGTELTGHLVPHDGLRWVCESSEAHIPLPAPAECCPMSCRLLGIG 120  
 QY 121 GHLSPGKLTLPBEPHCLKAPRLRCQSCCLAPGLRAGERSPESSLGRRRRNVNTAQ 180  
 DB 121 GHLSPGKLTLPBEPHCLKAPRLRCQSCCLAPGLRAGERSPESSLGRRRRNVNTAQ 180  
 QY 181 FQPPSTQATVPENOPAGTFVASTRAIDPEDEGAGGLRYTMDLPFSRSNQFSLDPVTGA 240  
 DB 181 FQPPSTQATVPENOPAGTFVASTRAIDPEDEGAGGLRYTMDLPFSRSNQFSLDPVTGA 240

241 VTTAELEBRETSTVFRVTLADHGMKPRSSALATLTITVTDTNDHPVFEQETYSURE 300  
Db VTTAELEBRETSTVFRVTLADHGMKPRSSALATLTITVTDTNDHPVFEQETYSURE 300  
Qy NLEVEYEVLTATAGDAPPNANILYRLLEGSGSPSEVEFIDPRSGVTRTGPPVREY 360  
Db NLEVEYEVLTATAGDAPPNANILYRLLEGSGSPSEVEFIDPRSGVTRTGPPVREY 360  
Qy ESYQUTVEASDQGRDGPGRSTTAAVFLSYEDDNDNAPQSEKRYVVOVBEDVTPGAPYLK 420  
Db ESYQUTVEASDQGRDGPGRSTTAAVFLSYEDDNDNAPQSEKRYVVOVBEDVTPGAPYLK 420  
Qy VTAASRDGSAVWHYSIMSNGARQFYLDAQTGLDVVSPLDYETTKYTLRVAQOGG 480  
Db VTAASRDGSAVWHYSIMSNGARQFYLDAQTGLDVVSPLDYETTKYTLRVAQOGG 480  
Qy RPLPSNBSGLVTYQVLINDNAPIFVSPFOATVLESVPLGYLVHVOALIDADAGDNRL 540  
Db RPLPSNBSGLVTYQVLINDNAPIFVSPFOATVLESVPLGYLVHVOALIDADAGDNRL 540  
Qy EYRLAGVGHDPFTJINNGTGMISVAEELDRBEVDYSGVEARDHGTPLTASASVTV 600  
Db EYRLAGVGHDPFTJINNGTGMISVAEELDRBEVDYSGVEARDHGTPLTASASVTV 600  
Qy LDVNDNNPTFPOPEYTVRLNEDAAVGTSVYVISAVDRAHSTYITQITSGNTRNFSTTS 660  
Db LDVNDNNPTFPOPEYTVRLNEDAAVGTSVYVISAVDRAHSTYITQITSGNTRNFSTTS 660  
Qy QSGGGLVSLALPLDYKLEROYLAVTASDGTRODPAQIYVNTDANTHRPVQSSHVTN 720  
Db QSGGGLVSLALPLDYKLEROYLAVTASDGTRODPAQIYVNTDANTHRPVQSSHVTN 720  
Qy VNEBDRPACTVVLISATDEDTGENARITYFMEDSIPOFRIDADTGAVTQAELEDYDQVS 780  
Db VNEBDRPACTVVLISATDEDTGENARITYFMEDSIPOFRIDADTGAVTQAELEDYDQVS 780  
Qy YTLAATAANDNGIPQSDPTTYLEILVNDVNDNAPOTLRBSYQGSYVEDVPPFTSVQISAT 840  
Db YTLAATAANDNGIPQSDPTTYLEILVNDVNDNAPOTLRBSYQGSYVEDVPPFTSVQISAT 840  
Qy DRDSGLANGRVFTFGGDDGDDGFIVESTSGIVRTLRRLDRBNVAQVYLRAYAVDKGMP 900  
Db DRDSGLANGRVFTFGGDDGDDGFIVESTSGIVRTLRRLDRBNVAQVYLRAYAVDKGMP 900  
Qy ARTPEMTVTYVLVDVNDNPPVFEQDEPDVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
Db ARTPEMTVTYVLVDVNDNPPVFEQDEPDVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
Qy VEGNIPEVFOQDIFSGELTALVDLYEDRPEVYVIOATSAVLBRATVHRLDRNDP 1020  
Db VEGNIPEVFOQDIFSGELTALVDLYEDRPEVYVIOATSAVLBRATVHRLDRNDP 1020  
Qy PVLGAFELLFNNTYNNRSSSPFGAIGRVPAHDPOISBLTYSFPRGNELSLVILNASTG 1080  
Db PVLGAFELLFNNTYNNRSSSPFGAIGRVPAHDPOISBLTYSFPRGNELSLVILNASTG 1080  
Qy EIKLSRALDNNRPLEAIVSVDSVHSTVQAQALRVITIDEMULTHSITRLLEDMSPER 1140  
Db EIKLSRALDNNRPLEAIVSVDSVHSTVQAQALRVITIDEMULTHSITRLLEDMSPER 1140  
Qy FLSPILGFIQAVATLTATPRPHVVVFNQVQRTDAPGCHILNVSISVQCPFGGPGPPL 1200  
Db FLSPILGFIQAVATLTATPRPHVVVFNQVQRTDAPGCHILNVSISVQCPFGGPGPPL 1200  
Qy PSEDLOERLYNRSILTAISAQVLPFDNICTAREPCENYMCVILRFDSSAPRTASS 1260  
Db PSEDLOERLYNRSILTAISAQVLPFDNICTAREPCENYMCVILRFDSSAPRTASS 1260  
Qy VLFRIPIHVGGLRCRCPPGFTGDCYETEVDLCYSRPCGPHGRCSRREGGYTCLCADGYTG 1320  
Db VLFRIPIHVGGLRCRCPPGFTGDCYETEVDLCYSRPCGPHGRCSRREGGYTCLCADGYTG 1320  
Qy BHCEVSASASGRCTPGVCNKGCTCVNLLVGGFKCDPCSDGFEKPYCOVTTTRSPASFITP 1380

1321 BHCEVSASASGRCTPGVCNKGCTCVNLLVGGFKCDPCSDGFEKPYCOVTTTRSPASFITP 1380  
Qy RGLRORFFHTALFATBERDGLLYNGRPNKBDPVALVEYIOEVOVLTFSAGESTTVS 1440  
Db RGLRORFFHTALFATBERDGLLYNGRPNKBDPVALVEYIOEVOVLTFSAGESTTVS 1440  
Qy PFVPGVSDGQWHTVQLKYNNKPLLQGTGLPQGSBOKVAVVYDGCCTGVALARGSVILG 1500  
Db PFVPGVSDGQWHTVQLKYNNKPLLQGTGLPQGSBOKVAVVYDGCCTGVALARGSVILG 1500  
Qy NYSCAAGTQGSXKSIDLTPPLLLGVPDLPESEFPVPMROFVGCNRLQVDSRHIDMAD 1560  
Db NYSCAAGTQGSXKSIDLTPPLLLGVPDLPESEFPVPMROFVGCNRLQVDSRHIDMAD 1560  
Qy FIANNGTVPCCPAKKNVCDNSTCANGGTCVNQOMAFSCCEPLGFGGKSCAQBMANPOHFL 1620  
Db FIANNGTVPCCPAKKNVCDNSTCANGGTCVNQOMAFSCCEPLGFGGKSCAQBMANPOHFL 1620  
Qy GSSLVAMHGLSLPISEPWTLSLMFRTRQADGVLLQALTRGRSTITLQLRBGHMLSVBGT 1680  
Db GSSLVAMHGLSLPISEPWTLSLMFRTRQADGVLLQALTRGRSTITLQLRBGHMLSVBGT 1680  
Qy GLQASSLRLEBRANDGDMHHAQALGASGGRGHAIIISFDYGOORABENLCPRIHGLHS 1740  
Db GLQASSLRLEBRANDGDMHHAQALGASGGRGHAIIISFDYGOORABENLCPRIHGLHS 1740  
Qy NITVGGIPGPAGVARGRGCLQGVRSVDPREGNSLDPHSGESINVEQGSLLDPCDSN 1800  
Db NITVGGIPGPAGVARGRGCLQGVRSVDPREGNSLDPHSGESINVEQGSLLDPCDSN 1800  
Qy PCPANSYCSNDMDSYSGSCPFGYDGNCTNVCNLPCEHOSVCTRKSPAPRGYCECPN 1860  
Db PCPANSYCSNDMDSYSGSCPFGYDGNCTNVCNLPCEHOSVCTRKSPAPRGYCECPN 1860  
Qy YLGPYCETRLIDQPCPRGMWGHPTGFCNCVSKSPEDDCKNTSBRCKEKNHYPPPGPT 1920  
Db YLGPYCETRLIDQPCPRGMWGHPTGFCNCVSKSPEDDCKNTSBRCKEKNHYPPPGPT 1920  
Qy CLLCDCTPFGSLSVCPEDBQCCPKKGVIGROCDRCNPFAYVYTNNGCEVNYSCPAI 1980  
Db CLLCDCTPFGSLSVCPEDBQCCPKKGVIGROCDRCNPFAYVYTNNGCEVNYSCPAI 1980  
Qy BAGIWMPTREFGDPAAPCPKGSFGTAVRHCDERGMPLPMLFNCTSTTFSELKGAERL 2040  
Db BAGIWMPTREFGDPAAPCPKGSFGTAVRHCDERGMPLPMLFNCTSTTFSELKGAERL 2040  
Qy QRNESGLDGSROQDALLRRAVTOHTAGYRGSVYKAVYQALATRLLAHSTRGGLSATQ 2100  
Db QRNESGLDGSROQDALLRRAVTOHTAGYRGSVYKAVYQALATRLLAHSTRGGLSATQ 2100  
Qy DVHFTENILRVGSLALDTANKRHELIQOTEGGTAMLIQHYEAVASALQAMRHTVLSPF 2160  
Db DVHFTENILRVGSLALDTANKRHELIQOTEGGTAMLIQHYEAVASALQAMRHTVLSPF 2160  
Qy TITVFNIVISVRLDKGNFAGAKLPRYEALRGEOPDLETTVIIPESVFRETPPVVRBAG 2220  
Db TITVFNIVISVRLDKGNFAGAKLPRYEALRGEOPDLETTVIIPESVFRETPPVVRBAG 2220  
Qy PGEAQBPEBELARRRHEPILSOGRAVASVITRYTLAGILPNNYDPDRKSLRVPKRPINT 2280  
Db PGEAQBPEBELARRRHEPILSOGRAVASVITRYTLAGILPNNYDPDRKSLRVPKRPINT 2280  
Qy PVASISVHDBELLPRALDKPVYQFRLLEBRTKICVPMNHSILVSGTGSAGSAGE 2340  
Db PVASISVHDBELLPRALDKPVYQFRLLEBRTKICVPMNHSILVSGTGSAGSAGE 2340  
Qy VVFNESHVSQCNHMTSPAVLMDVSRENGEIIPLKTLTYVALGVTLAALLTFPFLTL 2400  
Db VVFNESHVSQCNHMTSPAVLMDVSRENGEIIPLKTLTYVALGVTLAALLTFPFLTL 2400  
Qy LRIIRSNOHGRNMLTALGIAQLVPLGINQADLPACTVIALLLFLVYCTESMLLE 2460

```

Db 2401 LRLRSNONGIRRNLTAAQLAQLVFLGLINQADLPFACTIVAILLFLYLCTFSWALLE 2460
Qy 2461 ALHLVRLATVRDVNTGPMKFTYMLGNGVPAFTGLAVGLDPKGNPDCMTSTYDTLI 2520
Db 2461 ALHLVRLATVRDVNTGPMKFTYMLGNGVPAFTGLAVGLDPKGNPDCMTSTYDTLI 2520
Qy 2521 WSPGAVPAVAVNSVFLYTLAABASCAAOQGFEEKGPGVGLQPSFAVLLLSATWLLAL 2580
Db 2521 WSPGAVPAVAVNSVFLYTLAABASCAAOQGFEEKGPGVGLQPSFAVLLLSATWLLAL 2580
Qy 2581 LSVNSTTLPHYLPAFCNCICQGFPIFLSYLVLSKVRKAKLACSKRPSPDPAITTKSTL 2640
Db 2581 LSVNSTTLPHYLPAFCNCICQGFPIFLSYLVLSKVRKAKLACSKRPSPDPAITTKSTL 2640
Qy 2641 TSSYNCPSFYADRLQVYVDSAGSLHSTRSKSGSPSYPLRLRBSALNPGQPGG 2700
Db 2641 TSSYNCPSFYADRLQVYVDSAGSLHSTRSKSGSPSYPLRLRBSALNPGQPGG 2700
Qy 2701 DPGSLFLGQDQDHPDPTSDSLSDDDQSGSYASTHSDSEBEBEBEBEBAAPFGEQ 2760
Db 2701 DPGSLFLGQDQDHPDPTSDSLSDDDQSGSYASTHSDSEBEBEBEBEBAAPFGEQ 2760
Qy 2761 WBSLLPQGAERLPLHSTPQDGGPGKAPVPGDFTTAKSSGNGAPBERLRNGDALSR 2820
Db 2761 WBSLLPQGAERLPLHSTPQDGGPGKAPVPGDFTTAKSSGNGAPBERLRNGDALSR 2820
Qy 2821 EGSLLGPLPGSAOPHNGILKKCLPTITSEKSLRLPLRQCTGSSGSSASBGRGPP 2880
Db 2821 EGSLLGPLPGSAOPHNGILKKCLPTITSEKSLRLPLRQCTGSSGSSASBGRGPP 2880
Qy 2881 RPPPRQSLQEQNLGVNPIAMISIRKTVDEDSGSEFLFFNPLH 2923
Db 2881 RPPPRQSLQEQNLGVNPIAMISIRKTVDEDSGSEFLFFNPLH 2923

RESULT 2
CLR2 RAT STANDARD; PRT; 2144 AA.
AC 090YF2;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caderlin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal
DE growth factor-like domains 3) (Fragment).
GN CELSR2 OR MEGR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. High expression in
CC cerebellum and olfactory bulb. Weaker expression in cerebral
CC cortex, hippocampus and brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011529; BAA8687.1; -.
DR HSSP; P00740; 1EDM.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0004930; P:G-protein coupled receptor activity; ISS.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Caderlin.
DR InterPro; IPR001881; EGF Ca..
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormu_receptor.
DR InterPro; IPR002049; Laminin_Gf.
DR InterPro; IPR001791; Laminin_Gf.
DR pfam; PF00002; 7tm_2; 1. _cys_rich.
DR pfam; PF00028; caderlin; 3.
DR pfam; PF00008; EGF; 6.
DR pfam; PF01825; GPS; 1.
DR pfam; PF02793; HRM; 1.
DR pfam; PF00053; laminin_EGF; 1.
DR pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 3.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hornr; 1.
DR SMART; SM00282; Lamg; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS02068; CADHERIN_2; 4.
DR PROSITE; PS00221; GPS; 1.
DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE; PS00227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE; PS00261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE; PS01248; LAMININ_TYR_EGF; 1.
KM G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation.
FT 1 1605 EXTRACELLULAR (POTENTIAL).
FT 1 1626 1 (POTENTIAL).
FT 1 1626 CYTOPLASMIC (POTENTIAL).
FT 1 1641 2 (POTENTIAL).
FT 1 1641 EXTRACELLULAR (POTENTIAL).
FT 1 1663 3 (POTENTIAL).
FT 1 1663 CYTOPLASMIC (POTENTIAL).
FT 1 1684 4 (POTENTIAL).
FT 1 1705 4 (POTENTIAL).
FT 1 1706 EXTRACELLULAR (POTENTIAL).
FT 1 1726 5 (POTENTIAL).
FT 1 1744 5 (POTENTIAL).
FT 1 1744 CYTOPLASMIC (POTENTIAL).
FT 1 1765 6 (POTENTIAL).
FT 1 1789 6 (POTENTIAL).
FT 1 1810 EXTRACELLULAR (POTENTIAL).
FT 1 1816 7 (POTENTIAL).
FT 1 1837 7 (POTENTIAL).
FT 1 1838 CYTOPLASMIC (POTENTIAL).
FT 1 2144 CADHERIN 1.
FT 1 40 CADHERIN 2.
FT 1 146 CADHERIN 3.
FT 1 248 CADHERIN 4.
FT 1 371 CADHERIN 4.
FT 1 453 EGF-LIKE 1.
FT 1 511 EGF-LIKE 2.
FT 1 549 EGF-LIKE 3.
FT 1 591 CALCIUM-BINDING.
FT 1 592 LAMININ G-LIKE 1.

```



FT	DOMAIN	799	835	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	839	1016	LAMININ G-LIKE 2.
FT	DOMAIN	1018	1053	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1054	1092	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	1108	1147	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	1148	1180	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	1541	1593	GPS.
FT	DOMAIN	1968	1973	POLY-GLU.
FT	DISULFID	457	468	BY SIMILARITY.
FT	DISULFID	462	499	BY SIMILARITY.
FT	DISULFID	501	510	BY SIMILARITY.
FT	DISULFID	517	528	BY SIMILARITY.
FT	DISULFID	522	537	BY SIMILARITY.
FT	DISULFID	539	548	BY SIMILARITY.
FT	DISULFID	557	568	BY SIMILARITY.
FT	DISULFID	562	578	BY SIMILARITY.
FT	DISULFID	580	590	BY SIMILARITY.
FT	DISULFID	803	814	BY SIMILARITY.
FT	DISULFID	808	823	BY SIMILARITY.
FT	DISULFID	825	834	BY SIMILARITY.
FT	DISULFID	1022	1033	BY SIMILARITY.
FT	DISULFID	1027	1042	BY SIMILARITY.
FT	DISULFID	1044	1053	BY SIMILARITY.
FT	DISULFID	1057	1068	BY SIMILARITY.
FT	DISULFID	1062	1080	BY SIMILARITY.
FT	DISULFID	1082	1091	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1114	1131	BY SIMILARITY.
FT	DISULFID	1133	1146	BY SIMILARITY.
FT	DISULFID	1149	1161	BY SIMILARITY.
FT	DISULFID	1151	1168	BY SIMILARITY.
FT	DISULFID	1170	1179	BY SIMILARITY.
FT	MOD_RES	816	816	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	1035	1035	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	407	407	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	726	726	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	966	966	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1052	1052	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1125	1125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1249	1249	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1268	1268	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1286	1286	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1548	1548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1570	1570	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	2144 AA;	233480 MW;	6EA998C1BA655DECA CRC64;

Query Match	187;	6.4%;	Score 187;	DB 1;	Length 2244;	
Best Local Similarity	100.0%;		Pred. No. 3e-134;			
Matches	187;	Conservative	0;	Mismatches	0;	
			Indels	0;	Gaps	0;
Qy	1115	LRVTTITDDEMLTHSTITLRLIEDMSPERFLSPSLGLFIQAVATLATPDPHVVFVNVQDRTD	1174			
Db	340	LRVTTITDDEMLTHSTITLRLIEDMSPERFLSPSLGLFIQAVATLATPDPHVVFVNVQDRTD	399			
Qy	1175	APGGHILNVSLSVGQPPGPGGAPPPLPSEBDLOERLYLNRSLTATSAORVLPFDNITCLR	1233			
Db	400	APGGHILNVSLSVGQPPGPGGAPPPLPSEBDLOERLYLNRSLTATSAORVLPFDNITCLR	459			
Qy	1235	EPCENNYMCVSVLAPDSSAPFIASSSVLFRPIHPVGLRCRCPPGFTGDYCTEVDLCYS	1294			
Db	460	EPCENNYMCVSVLAPDSSAPFIASSSVLFRPIHPVGLRCRCPPGFTGDYCTEVDLCYS	519			
Qy	1295	RPCGPHG 1301				
Db	520	RPCGPHG 526				

RESULT 3  
CLR2\_MOUSE

ID AC CLR2 MOUSE STANDARD; PRT: 2920 AA.  
DT 09R0M0\_099K26\_092Z234;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Cdherrin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)  
DB (Ensembl).  
GN CELSR2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxId:10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99418630; PubMed=10490098;  
RA Ueki T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,  
RA Takeichi M., Demura T.;  
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell  
RL polarity under the control of frizzled."  
RL Cell 98:585-595(1999).  
RN [2]  
RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.  
RX MEDLINE=20253755; PubMed=10790539;  
RA Forrester C.J., Barclay J., Rees M., Little P.P.R.;  
RT "Chromosomal localization of Celser2 and Celser3 in the mouse; Celser3 is  
RL a candidate for the tippy (tip) lethal mutant on chromosome 9."  
RN [3]  
RP SEQUENCE OF 2014-2920 FROM N.A.  
RX TISSUE=Breast tumor;  
RL MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Matusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.U., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McLean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP DEVELOPMENTAL STAGE.  
RX MEDLINE=21899555; PubMed=11850187;  
RA Tisserat F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;  
RT "Developmental expression profiles of Celser (Flamingo) genes in the  
RT mouse."  
RL Mech. Dev. 112:157-160(2002).  
CC -1- FUNCTION: Receptor that may have an important role in cell/cell  
CC signaling during nervous system formation.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.  
CC -1- DEVELOPMENTAL STAGE: Predominantly expressed in the developing  
CC CNS, the emerging dorsal root ganglia and cranial ganglia. In the  
CC CNS, expression is uniform along the rostrocaudal axis. During  
CC gastrulation, it is expressed within the anterior neural ectoderm.  
CC At E10, expression is strong in the ventricular zones (VZ) in all  
CC sectors of the brain, and lower in the marginal zones (MZ) in all  
CC between E12 and E15, expression is prominent in the brain. It is  
CC strong in VZ, lower in MZ, except in telecephalic MZ where it is  
CC predominant. The intensity is higher in all VZ, and lower in  
CC differentiating fields than in VZ, except in the cerebral  
CC hemispheres, and to a lesser extent in the tectum and cerebellum.

A weak expression is also observed in the fetal lungs, kidney and epithelia. In the newborn and postnatal stages, expression remains restricted to the VZ as well as in migrating and postmigratory cells throughout the brain.

-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

-1- SIMILARITY: Contains 9 cadherin domains.

-1- SIMILARITY: Contains 8 EGF-like domains.

-1- SIMILARITY: Contains 2 laminin G-like domains.

-1- SIMILARITY: Contains 1 laminin EGF-like domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AB028499; BAA84070.1; -

EMBL; AF031573; AAC68837.1; -

EMBL; BC005499; AA05499.1; -

HSSP; P00740; 1EDM.

MCD; MGI:1658235; Celarg2.

GO; GO:0016021; C: integral to membrane; ISS.

GO; GO:0004930; F: G-protein coupled receptor activity; ISS.

InterPro; IPR0010152; Asx\_hydroxyl.

InterPro; IPR002126; Cadherin.

InterPro; IPR006209; EGF\_like.

InterPro; IPR000832; GPCR\_secretin.

InterPro; IPR001879; hormn\_receptor.

InterPro; IPR002049; Laminin\_EGF.

InterPro; IPR001791; Laminin\_G.

InterPro; IPR002003; PKD\_cys\_rich.

Pfam; PF00002; 7tm\_2; 1.

Pfam; PF00028; cadherin; 9.

Pfam; PF00008; EGF; 5.

Pfam; PF01825; GPS; 1.

Pfam; PF02793; HRM; 1.

Pfam; PF00053; Laminin\_EGF; 1.

Pfam; PF00054; Laminin\_G; 2.

PRINTS; PRO0205; CADHERIN.

PRINTS; PRO0011; EGF\_LAMININ.

PRINTS; PRO0249; GPCR\_SECRETIN.

SMART; SM00112; CA; 9.

SMART; SM00180; EGF\_Lam; 1.

SMART; SM00303; GPS; 1.

SMART; SM00008; Hormr; 1.

SMART; SM00282; LamG; 2.

PROSITE; PS00010; ASX\_HYDROXYL; 1.

PROSITE; PS00232; CADHERIN\_1; 6.

PROSITE; PS00268; CADHERIN\_2; 9.

PROSITE; PS00022; EGF\_1; 6.

PROSITE; PS01186; EGF\_2; 2.

PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.

PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; FALSE\_NEG.

PROSITE; PS00227; G\_PROTEIN\_RECP\_F2\_3; 1.

PROSITE; PS00264; G\_PROTEIN\_RECP\_F2\_4; 1.

PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.

PROSITE; PS00221; GPS; 1.

PROSITE; PS00025; LAM\_G\_DOMAIN; 2.

G-protein coupled receptor; Transmembrane; Glycoprotein; EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat; Developmental protein; Hydroxylation; Signal.

CHAIN 1 31 POTENTIAL.

FT CHAIN 32 2920 CADHERIN EGF LAG SEVEN-PASS G-TYPE

FT DOMAIN 32 2381 RECEPTOR 2.

FT TRANSMEM 2382 2402 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 2403 2414 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 2415 2434 2 (POTENTIAL).

FT DOMAIN 2435 2439 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	2440	2460	3 (POTENTIAL).
FT	DOMAIN	2461	2481	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2482	2502	4 (POTENTIAL).
FT	DOMAIN	2503	2519	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2520	2540	5 (POTENTIAL).
FT	DOMAIN	2541	2564	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2565	2585	6 (POTENTIAL).
FT	DOMAIN	2586	2592	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2593	2613	7 (POTENTIAL).
FT	DOMAIN	2614	2920	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	182	289	CADHERIN 1.
FT	DOMAIN	290	339	CADHERIN 2.
FT	DOMAIN	400	506	CADHERIN 3.
FT	DOMAIN	507	611	CADHERIN 4.
FT	DOMAIN	612	713	CADHERIN 5.
FT	DOMAIN	714	816	CADHERIN 6.
FT	DOMAIN	817	922	CADHERIN 7.
FT	DOMAIN	923	1024	CADHERIN 8.
FT	DOMAIN	1029	1147	CADHERIN 9.
FT	DOMAIN	1229	1287	EGF-LIKE 1 (ATYPICAL).
FT	DOMAIN	1289	1319	EGF-LIKE 2, CALCIUM-BINDING.
FT	DOMAIN	1329	1367	EGF-LIKE 3, CALCIUM-BINDING.
FT	DOMAIN	1368	1572	LAMININ G-LIKE 1.
FT	DOMAIN	1575	1611	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	1615	1792	EGF-LIKE 5, CALCIUM-BINDING.
FT	DOMAIN	1788	1830	EGF-LIKE 6, CALCIUM-BINDING.
FT	DOMAIN	1831	1868	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	1864	1923	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	1924	1956	GPS.
FT	DOMAIN	2317	2369	POLY-GLU.
FT	DOMAIN	2744	2749	BY SIMILARITY.
FT	DISULFID	1293	1304	BY SIMILARITY.
FT	DISULFID	1298	1313	BY SIMILARITY.
FT	DISULFID	1315	1318	BY SIMILARITY.
FT	DISULFID	1333	1344	BY SIMILARITY.
FT	DISULFID	1338	1354	BY SIMILARITY.
FT	DISULFID	1356	1366	BY SIMILARITY.
FT	DISULFID	1579	1590	BY SIMILARITY.
FT	DISULFID	1584	1599	BY SIMILARITY.
FT	DISULFID	1601	1610	BY SIMILARITY.
FT	DISULFID	1792	1803	BY SIMILARITY.
FT	DISULFID	1798	1818	BY SIMILARITY.
FT	DISULFID	1820	1829	BY SIMILARITY.
FT	DISULFID	1833	1844	BY SIMILARITY.
FT	DISULFID	1838	1856	BY SIMILARITY.
FT	DISULFID	1858	1867	BY SIMILARITY.
FT	DISULFID	1868	1900	BY SIMILARITY.
FT	DISULFID	1880	1907	BY SIMILARITY.
FT	DISULFID	1909	1922	BY SIMILARITY.
FT	DISULFID	1925	1937	BY SIMILARITY.
FT	DISULFID	1927	1944	BY SIMILARITY.
FT	DISULFID	1946	1955	BY SIMILARITY.
FT	MOD RES	1592	1592	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	486	486	N-LINKED (GLCNAC... ) (POTENTIAL).

Query Match 5.7%; Score 168; DB 1; Length 2920;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-119;  
 Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	915	NNNPVPEDEDEPVEENSP	IGLAVARTADPDGNTAQM	QIVESGNIPEVOLDIF	974
DB	916	NNNPVFEDEDEPVEENSP <td>IGLAVARTADPDGNTAQM <td>QIVESGNIPEVOLDIF <td>975</td> </td></td>	IGLAVARTADPDGNTAQM <td>QIVESGNIPEVOLDIF <td>975</td> </td>	QIVESGNIPEVOLDIF <td>975</td>	975
QY	975	SGELTALVDLDYEDRPEYLV <td>IQATSAPLVSRATYVRL <td>LDNDNPVYLGNEFILLNNV</td> <td>1034</td> </td>	IQATSAPLVSRATYVRL <td>LDNDNPVYLGNEFILLNNV</td> <td>1034</td>	LDNDNPVYLGNEFILLNNV	1034
DB	976	SGELTALVDLDYEDRPEYLV <td>IQATSAPLVSRATYVRL <td>LDNDNPVYLGNEFILLNNV</td> <td>1035</td> </td>	IQATSAPLVSRATYVRL <td>LDNDNPVYLGNEFILLNNV</td> <td>1035</td>	LDNDNPVYLGNEFILLNNV	1035
QY	1035	TRRSSFGGATGRVPADPD	ISDLTTSFERGNEISLV <td>LAASGTEL</td> <td>1082</td>	LAASGTEL	1082
DB	1036	TRRSSFGGATGRVPADPD	ISDLTTSFERGNEISLV <td>LAASGTEL</td> <td>1083</td>	LAASGTEL	1083

RESULT 4

ID	CLRT_MOUSE	STANDARD;	PRT;	3034 AA.
AC	O35161.			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cadherin EGF LAG seven-pass G-type receptor 1 precursor.			
GN	CELSR1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEIDLIN=99077689; PubMed=98586937;			
RA	Hadjantonakis A.-K., Formstone C.J., Little P.F.R.;			
RA	"Celsr1, a neutral-specific gene encoding an unusual seven-pass			
RT	transmembrane receptor, maps to mouse chromosome 15 and human			
RL	chromosome 22qter."			
RL	Genomics 45:97-104(1997).			
RN	[3]			
RP	TISSUE SPECIFICITY.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RX	MEIDLIN=97480720; PubMed=9339365;			
RA	Hadjantonakis A.-K., Sheward W.J., Harmar A.J., de Galan L.,			
RA	Hoovers J.M.N., Little P.F.R.;			
RT	"Celsr1, a neutral-specific gene encoding an unusual seven-pass			
RT	transmembrane receptor, maps to mouse chromosome 15 and human			
RL	chromosome 22qter."			
RL	Genomics 45:97-104(1997).			
RN	[3]			
RP	DEVELOPMENTAL STAGE.			
RX	MEIDLIN=21839555; PubMed=11850187;			
RA	Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;			
RT	"Developmental expression profiles of Celsr (Flamingo) genes in the			
RT	mouse."			
RL	Mech. Dev. 112:157-160(2002).			
CC	-1- FUNCTION: Receptor that may have an important role in cell/cell			
CC	signaling during nervous system formation.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- TISSUE SPECIFICITY: Expressed in the brain, where it is localized			
CC	principally in the ependymal cell layer, choroid plexus and the			
CC	area postrema. Also found in spinal chord and in the eye.			
CC	-1- DEVELOPMENTAL STAGE: First detected at E6. Predominantly expressed			
CC	in the developing CNS, the emerging dorsal root ganglia and			
CC	cranial ganglia. In the CNS, expression is uniform along the			
CC	rostrocaudal axis. During gastrulation, it is expressed in the			
CC	vicinity of the primitive streak, and becomes predominant in that			
CC	area at late gastrulation. At E10, detected in ventricular zones			
CC	(VZ), but not in marginal zones (MZ), and weakly in other			
CC	structures. Between E12 and E15, a high expression is present in			
CC	the VZ in all brain areas. No expression in differentiated			
CC	neuronal fields. In the newborn and postnatal stages, expression			
CC	remains restricted to the VZ. Also found weakly in fetal lungs,			
CC	kidney and epithelia.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-1- SIMILARITY: Contains 9 cadherin domains.			
CC	-1- SIMILARITY: Contains 8 EGF-like domains.			
CC	-1- SIMILARITY: Contains 2 laminin G-like domains.			
CC	-1- SIMILARITY: Contains 1 laminin EGF-like domain.			
CC	-1- SIMILARITY: Contains 1 GPS domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF031572; AAC68836.1; -			
DR	PIR; T14119; T14119.			
DR	MED; MG1:1100883; Celsr1.			
DR	HSSP; P00749; IURK.			

DR	InterPro: IPR000152; Asx hydroxylase.
DR	InterPro: IPR002126; Cadherin.
DR	InterPro: IPR006209; EGF-like.
DR	InterPro: IPR000832; GPCR secretin.
DR	InterPro: IPR001879; horum_receptor.
DR	InterPro: IPR002049; laminin_egg.
DR	InterPro: IPR001791; laminin G.
DR	InterPro: IPR000203; ptd_cys_rich.
DR	pfam: PF00002; 7tm_2; 1_pcd_rich.
DR	pfam: PF00028; cadherin; 9.
DR	pfam: PF00008; EGF; 6.
DR	pfam: PF01825; GPS; 1.
DR	pfam: PF02793; HRM; 1.
DR	pfam: PF00053; laminin EGF; 1.
DR	pfam: PF00054; laminin G; 1.
DR	PRINTS: PR00205; CADHERIN.
DR	PRINTS: PR00011; EGF_LAMININ.
DR	PRINTS: PR00249; GPCRSECRETIN.
DR	SMART: SM00112; CA; 9.
DR	SMART: SM00180; EGF_lam; 1.
DR	SMART: SM00303; GPS; 1.
DR	SMART: SM00008; Homr; 1.
DR	SMART: SM00282; Lamg; 2.
DR	PROSITE: PS00010; ASX_HYDROXYL; 2.
DR	PROSITE: PS00232; CADHERIN_1; 7.
DR	PROSITE: PS02688; CADHERIN_2; 9.
DR	PROSITE: PS00023; EGF_1; 6.
DR	PROSITE: PS01166; EGF_2; 2.
DR	PROSITE: PS02021; GPS; 1.
DR	PROSITE: PS00649; G_PROTEIN_RECP_F2_1; FALSE_NEG.
DR	PROSITE: PS00650; G_PROTEIN_RECP_F2_2; FALSE_NEG.
DR	PROSITE: PS02027; G_PROTEIN_RECP_F2_3; 1.
DR	PROSITE: PS02026; G_PROTEIN_RECP_F2_4; 1.
DR	PROSITE: PS01248; LAMININ_TYRE_EGF; 1.
DR	PROSITE: PS00025; LAM_G_DOMAIN; 2.
KM	G-protein coupled receptor; transmembrane; glycoprotein;
KM	EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat
KM	Developmental protein; Hydroxylation; Signal.
KT	SIGNAL
FT	1
FT	29
FT	30
FT	3034
FT	DOMAIN
FT	21
FT	2484
FT	2485
FT	2505
FT	2506
FT	2516
FT	2517
FT	2537
FT	2538
FT	2542
FT	2543
FT	2563
FT	2564
FT	2587
FT	2588
FT	2589
FT	2609
FT	2625
FT	2626
FT	2646
FT	2647
FT	2670
FT	2671
FT	2691
FT	2692
FT	2694
FT	2715
FT	2716
FT	2717
FT	3034
FT	3034
FT	368
FT	474
FT	474
FT	369
FT	475
FT	580
FT	702
FT	581
FT	703
FT	804
FT	805
FT	907
FT	1014
FT	1015
FT	1116
FT	1121
FT	1239
FT	1318
FT	1376
FT	1378
FT	1414
FT	1418
FT	1456
FT	1457
FT	1661
FT	1664
FT	1700
FT	1704
FT	1885
FT	1922
FT	1923
FT	1961
FT	DOMAIN
FT	21
FT	2484
FT	2485
FT	2505
FT	2506
FT	2516
FT	2517
FT	2537
FT	2538
FT	2542
FT	2543
FT	2563
FT	2564
FT	2587
FT	2588
FT	2589
FT	2609
FT	2625
FT	2626
FT	2646
FT	2647
FT	2670
FT	2671
FT	2691
FT	2692
FT	2694
FT	2715
FT	2716
FT	3034
FT	3034
FT	368
FT	474
FT	474
FT	369
FT	475
FT	580
FT	702
FT	581
FT	703
FT	804
FT	805
FT	907
FT	1014
FT	1015
FT	1116
FT	1121
FT	1239
FT	1318
FT	1376
FT	1378
FT	1414
FT	1418
FT	1456
FT	1457
FT	1661
FT	1664
FT	1700
FT	1704
FT	1885
FT	1922
FT	1923
FT	1961
FT	DOMAIN
FT	21
FT	2484
FT	2485
FT	2505
FT	2506
FT	2516
FT	2517
FT	2537
FT	2538
FT	2542
FT	2543
FT	2563
FT	

```

hydroxyl.
in.
ke.
secretin.
receptor.
n_BGF.
n_G.
s_rich.
IN.
1.
XYL; 2.
1; 7.
2; 9.
RECEP_F2_1; FALSE_NEG.
RECEP_F2_2; FALSE_NEG.
RECEP_F2_3; 1.
RECEP_F2_4; 1.
TYPE_BGF; 1.
AIN; 2.
Transmembrane; Glycoprotein;
adding; laminin EGF-like domain
oxylation; Signal.
POTENTIAL.
CADERIN EGF LAG SEVEN-PASS
RECEPTOR 1.
EXTRACELLULAR (POTENTIAL) .
1 (POTENTIAL) .
CYTOPLASMIC (POTENTIAL) .
2 (POTENTIAL) .
EXTRACELLULAR (POTENTIAL) .
3 (POTENTIAL) .
CYTOPLASMIC (POTENTIAL) .
4 (POTENTIAL) .
EXTRACELLULAR (POTENTIAL) .
5 (POTENTIAL) .
CYTOPLASMIC (POTENTIAL) .
6 (POTENTIAL) .
EXTRACELLULAR (POTENTIAL) .
7 (POTENTIAL) .
CYTOPLASMIC (POTENTIAL) .
CADERIN 1.
CADERIN 2.
CADERIN 3.
CADERIN 4.
CADERIN 5.
CADERIN 6.
CADERIN 7.
CADERIN 8.
CADERIN 9.
EGF-LIKE 1, CALCIUM-BINDING
EGF-LIKE 2, CALCIUM-BINDING
EGF-LIKE 3, CALCIUM-BINDING
LAMININ G-LIKE 1
EGF-LIKE 4, CALCIUM-BINDING
LAMININ G-LIKE 2.
EGF-LIKE 5, CALCIUM-BINDING
EGF-LIKE 6, CALCIUM-BINDING

```

FT DOMAIN 1962 1994 EGF-LIKE 7, CALCIUM-BINDING.  
 FT DOMAIN 1996 2031 EGF-LIKE 8, CALCIUM-BINDING.  
 FT DOMAIN 2037 2070 LAMININ EGF-LIKE.  
 FT DOMAIN 2423 2475 GFS.  
 FT DOMAIN 2674 2678 POLY-LEU.  
 FT DISULFID 1322 1333 BY SIMILARITY.  
 FT DISULFID 1327 1364 BY SIMILARITY.  
 FT DISULFID 1366 1375 BY SIMILARITY.  
 FT DISULFID 1382 1393 BY SIMILARITY.  
 FT DISULFID 1387 1402 BY SIMILARITY.  
 FT DISULFID 1404 1413 BY SIMILARITY.  
 FT DISULFID 1422 1433 BY SIMILARITY.  
 FT DISULFID 1427 1443 BY SIMILARITY.  
 FT DISULFID 1445 1455 BY SIMILARITY.  
 FT DISULFID 1668 1679 BY SIMILARITY.  
 FT DISULFID 1673 1699 BY SIMILARITY.  
 FT DISULFID 1690 1699 BY SIMILARITY.  
 FT DISULFID 1891 1902 BY SIMILARITY.  
 FT DISULFID 1896 1911 BY SIMILARITY.  
 FT DISULFID 1913 1922 BY SIMILARITY.  
 FT DISULFID 1926 1937 BY SIMILARITY.  
 FT DISULFID 1931 1949 BY SIMILARITY.  
 FT DISULFID 1951 1960 BY SIMILARITY.  
 FT DISULFID 1968 1981 BY SIMILARITY.  
 FT DISULFID 1983 1993 BY SIMILARITY.  
 FT DISULFID 2000 2015 BY SIMILARITY.  
 FT DISULFID 2002 2018 BY SIMILARITY.  
 FT DISULFID 2020 2030 BY SIMILARITY.  
 FT MOD\_RES 1681 1681 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 1904 1904 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 236 236 N-LINKED (GLCNAC).  
 FT CARBOHYD 561 561 N-LINKED (GLCNAC).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC).  
 FT CARBOHYD 793 793 N-LINKED (GLCNAC).  
 FT CARBOHYD 1129 1129 N-LINKED (GLCNAC).  
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC).  
 FT CARBOHYD 1228 1228 N-LINKED (GLCNAC).  
 FT CARBOHYD 1264 1264 N-LINKED (GLCNAC).  
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC).  
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC).  
 FT CARBOHYD 1591 1591 N-LINKED (GLCNAC).  
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC).  
 FT CARBOHYD 1655 1655 N-LINKED (GLCNAC).  
 FT CARBOHYD 1994 1994 N-LINKED (GLCNAC).  
 FT CARBOHYD 2118 2118 N-LINKED (GLCNAC).  
 FT CARBOHYD 2137 2137 N-LINKED (GLCNAC).  
 FT CARBOHYD 2144 2144 N-LINKED (GLCNAC).  
 FT CARBOHYD 2155 2155 N-LINKED (GLCNAC).  
 FT CARBOHYD 2160 2160 N-LINKED (GLCNAC).  
 FT CARBOHYD 2272 2272 N-LINKED (GLCNAC).  
 FT CARBOHYD 2430 2430 N-LINKED (GLCNAC).  
 FT CARBOHYD 2452 2452 N-LINKED (GLCNAC).  
 FT CARBOHYD 2538 2538 N-LINKED (GLCNAC).  
 SQ SEQUENCE 3034 AA; 330477 MW; EEP38180AF5ED8A8 CMC64;

Query Match 0.8%; Score 23; DB 1; Length 3034;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-09;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1219 ISAGRTLPDNDTCAREPCNTM 1241  
 Db 1309 ISAGRTLPDNDTCAREPCNTM 1331

RESULT 5  
 CLRL\_HUMAN STANDARD; PRT; 3014 AA.  
 AC Q9NY06; Q95722; Q9BWS; Q9Y506; Q9Y526;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFm12).

GN CELSR1 OR CDHPS OR FM12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 OK [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2002599; PubMed=10716726;  
 RT "Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Brunkewich R., Beare D.M., Clamp M., Smink L.J., Amescough R., Almeida J.N., Beasley O.P., Bagguley C., Bailey J., Barlow K.P., Bates K.N., Burgess J., Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J., Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M., Cobley V.B., Cole C.G., Collier R.E., Connor R., Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E., Dharm P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G., Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C., Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A., Laird G.K., Langford C.F., Levertha M.A., Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T., Mcclay J., McLaren S., Mcmurray A.A., Milne S.A., Mortimore B.J., Odell C.N., Pavlit R., Pearce A.V., Pearson D., Phillimore B.J.C.T., Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Rose M.T., Scott C.B., Shira H.K., Skuce C.D., Smalley S., Smith M.L., Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wallis J.M., Whiteley M.N., Willey D.L., Williams L., Williams S.A., Williamson H., Wilner T.E., Wilting L., Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T., Dorman A., Fang F., Fan Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Lon P., Malek E., Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z., Zhan M., Zhang G., Chisese S., Murray J., Miller N., Minx P., Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saita S., Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow J.P., Bdeimann L., Kim U.J., Shitaya H., Simon M.I., Dumanaki J.B., Peyraud M., Kedra D., Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L., Tilahun Y., Wright H.;  
 "The DNA sequence of human chromosome 22."  
 RL Nature 402:489-495(1999).  
 RN [3]  
 RP SEQUENCE OF 624-3014 FROM N.A. (ISOPFORM 2).  
 RC TISSUE=Kidney;  
 RX MEDLINE=2388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hekman S., Garcia A.M., Gay L.J., Hultky S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

RA Fahay J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schin J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9NY06-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NY06-2; Sequence=VSP_002011, VSP_002012;
CC Note=No experimental confirmation available;
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi-sib.ch).
CC -----
DR EMBL; AF211024; AAF61930.1; -
DR EMBL; AL0211392; CAB50707.1; -
DR EMBL; AL031557; CAB45020.1; AUT_INIT.
DR EMBL; AL031588; CAB38413.1; -
DR EMBL; BC000059; AAH00059.1; -
DR Genew; HGNC:1850; CELSR1.
DR MIM; 604523; -
DR HSSP; P00749; IURK.
DR GO; GO:0016021; C:Integral to membrane, NAS.
DR GO; GO:0080014; P:calcium-dependent cell adhesion molecule ac. .; NAS.
DR GO; GO:0004930; P:G-protein coupled receptor activity; NAS.
DR GO; GO:0007339; P:neurogenesis; NAS.
DR InterPro; IPR000152; Aex hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000832; GPCR secretin.
DR InterPro; IPR001879; hormn receptor.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PRO_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 8.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF00008; GFS; 1.
DR Pfam; PF001825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00053; laminin_EGF_1.
DR Pfam; PF00054; laminin_G_1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF-LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SMO0112; CA_8.
DR SMART; SMO0180; EGF_Lam_1.
DR SMART; SMO0303; GPS; 1.
DR SMART; SMO0008; Hornr; 1.
DR SMART; SMO0282; Lamg; 2.
DR PROSITE; PS00610; ASX_HYDROXYL; 2.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS0268; CADHERIN_2; 9.
DR PROSITE; PS00022; EGF_1; 6.

```

[illegible]



```

FT DISULFID 1944 1959 BY SIMILARITY.
FT DISULFID 1961 1970 BY SIMILARITY.
FT DISULFID 1974 1985 BY SIMILARITY.
FT DISULFID 1979 1997 BY SIMILARITY.
FT DISULFID 1999 2008 BY SIMILARITY.
FT DISULFID 2016 2029 BY SIMILARITY.
FT DISULFID 2031 2041 BY SIMILARITY.
FT DISULFID 2048 2063 BY SIMILARITY.
FT DISULFID 2050 2066 BY SIMILARITY.
FT DISULFID 2068 2078 BY SIMILARITY.
FT MOD RES 1952 1952 HYDROXYLATION (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 2713 2713 L -> P (IN REF. 2).
FT CONFLICT 3024 3024 R -> P (IN REF. 2).
SQ SEQUENCE 3301 AA; 358455 MW; A6818F2DF7F4DEB6 CRC64;

```

```

Query Match 0.6%; Score 18; DB 1; Length 3301;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2495 GLAVGLDPEGYGNPDFCM 2512
Db 2649 GLAVGLDPEGYGNPDFCM 2666

```

```

RESULT 7
CLR3 HUMAN
AC 09NY07: 075092; STANDARD; PRT; 3312 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo
DE homolog 1) (hFml1) (Multiple epidermal growth factor-like domains 2)
DE (Epidermal growth factor-like 1).
GN CELSR3 OR CDHPL1 OR FMI1 OR EGF1 OR MEGF2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
RT feature of protocadherin genes";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RP [2]
RP SEQUENCE OF 1954-3312 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9836089; PubMed=9691030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

```

```

CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF231023; AAF61929.1; -.
CC EMBL; AB011536; BAA32464.1; -.
CC HSSP; P00740; IEDM.
CC Genew; HGNC:3230; CELSR3.
CC MTM; 604264; -.
DR GO; GO:0005198; F: structural molecule activity; NAS.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormu_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR001791; laminin_G.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G_2.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hornr; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS00022; EGF_2; 4.
DR PROSITE; PS01186; GPS; 1.
DR PROSITE; PS01221; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR PROSITE; PS50025; LAMININ_TYPE_EGF; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 32
FT CHAIN 33 3312
FT DOMAIN 33 2540
FT TRANSMEM 2541 2561
FT DOMAIN 2562 2572
FT TRANSMEM 2573 2593
FT DOMAIN 2594 2601
FT TRANSMEM 2602 2622
FT DOMAIN 2623 2643
FT TRANSMEM 2644 2664
FT DOMAIN 2665 2681
FT TRANSMEM 2682 2702

```



```

PT DOMAIN 2703 2725 CYTOPLASMIC (POTENTIAL).
PT TRANSSEM 2726 2746 6 (POTENTIAL).
PT TRANSSEM 2747 2753 EXTRACELLULAR (POTENTIAL).
PT TRANSSEM 2754 2774 7 (POTENTIAL).
PT TRANSSEM 2775 3312 CYTOPLASMIC (POTENTIAL).
PT DOMAIN 326 433 CADHERIN 1.
PT DOMAIN 434 545 CADHERIN 2.
PT DOMAIN 546 651 CADHERIN 3.
PT DOMAIN 652 756 CADHERIN 4.
PT DOMAIN 757 858 CADHERIN 5.
PT DOMAIN 859 961 CADHERIN 6.
PT DOMAIN 962 1067 CADHERIN 7.
PT DOMAIN 1068 1169 CADHERIN 8.
PT DOMAIN 1170 1265 CADHERIN 9.
PT DOMAIN 1266 1433 EGF-LIKE 1, CALCIUM-BINDING.
PT DOMAIN 1434 1471 EGF-LIKE 2, CALCIUM-BINDING.
PT DOMAIN 1475 1514 EGF-LIKE 3, CALCIUM-BINDING.
PT DOMAIN 1515 1719 LAMININ G-LIKE 1.
PT DOMAIN 1722 1758 EGF-LIKE 4, CALCIUM-BINDING.
PT DOMAIN 1764 1944 LAMININ G-LIKE 2.
PT DOMAIN 1946 1982 EGF-LIKE 5, CALCIUM-BINDING.
PT DOMAIN 1983 2020 EGF-LIKE 6, CALCIUM-BINDING.
PT DOMAIN 2021 2053 EGF-LIKE 7, CALCIUM-BINDING.
PT DOMAIN 2055 2090 EGF-LIKE 8, CALCIUM-BINDING.
PT DOMAIN 2096 2131 LAMININ EGF-LIKE.
PT DOMAIN 2477 2529 GPS.
PT DISULFID 1379 1390 BY SIMILARITY.
PT DISULFID 1384 1421 BY SIMILARITY.
PT DISULFID 1423 1432 BY SIMILARITY.
PT DISULFID 1439 1450 BY SIMILARITY.
PT DISULFID 1444 1459 BY SIMILARITY.
PT DISULFID 1461 1470 BY SIMILARITY.
PT DISULFID 1479 1490 BY SIMILARITY.
PT DISULFID 1484 1500 BY SIMILARITY.
PT DISULFID 1502 1513 BY SIMILARITY.
PT DISULFID 1726 1737 BY SIMILARITY.
PT DISULFID 1731 1746 BY SIMILARITY.
PT DISULFID 1748 1757 BY SIMILARITY.
PT DISULFID 1950 1961 BY SIMILARITY.
PT DISULFID 1955 1970 BY SIMILARITY.
PT DISULFID 1972 1981 BY SIMILARITY.
PT DISULFID 1985 1996 BY SIMILARITY.
PT DISULFID 1990 2008 BY SIMILARITY.
PT DISULFID 2010 2019 BY SIMILARITY.
PT DISULFID 2027 2040 BY SIMILARITY.
PT DISULFID 2042 2052 BY SIMILARITY.
PT DISULFID 2059 2074 BY SIMILARITY.
PT DISULFID 2061 2077 BY SIMILARITY.
PT DISULFID 2079 2089 BY SIMILARITY.
PT MOD RES 1963 1963 HYDROXYLATION (POTENTIAL).
PT CARBOHYD 632 632 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1222 1222 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1317 1317 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1327 1327 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1327 1327 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1449 1649 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1713 1713 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 1770 1770 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2053 2053 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2177 2177 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2386 2386 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2474 2474 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 2506 2506 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CONFLICT 2158 2158 G -> GLGAG (IN REF. 2).
SQ SEQUENCE 3312 AA; BEC208703651A4AS CRC64;

```

```

Query Match 0.63; Score 18; DB 1; Length 3312;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2495 GLAVGLDPGYNPFCM 2512

```

```

DB 2658 GLAVGLDPGYNPFCM 2675
RESULT 8
CLRS_RAT STANDARD; PRT; 3313 AA.
AC 088278;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Multiple
DE epidermal growth factor-like domains 2).
GN CELSR3 OR MEKFP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC cerebellum, olfactory bulb, cerebral cortex, hippocampus and
CC brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011528; BAA32459.1; -.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR000152; Abx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PKD_cye_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin_9.
DR Pfam; PF00008; EGF_6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G_1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGFLAMININ.
DR PRINTS; PR00243; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; HormR; 1.

```



```

DR EMBL; AB000895; BAA2133.1; -.
DR PIR; PC4297; PC4297.
DR Genew; HGNC:13681; PCDH16.
DR MTM; 603057; -.
DR GO; GO:0016539; P:calcium-dependent cell-cell adhesion; NMS.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 27.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 27.
DR PROSITE; PS00232; CADHERIN_1; 18.
DR PROSITE; PS50268; CADHERIN_2; 27.
DR KEGG; Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal; Multigene family.
FT SIGNAL 1 42
FT CHAIN 3 3298
FT DOMAIN 43 2940
FT TRANSFEM 2941 2961
FT TRANSFEM 2962 3298
FT DOMAIN 43 143
FT DOMAIN 144 255
FT DOMAIN 256 367
FT DOMAIN 367 472
FT DOMAIN 474 578
FT DOMAIN 579 685
FT DOMAIN 686 790
FT DOMAIN 791 894
FT DOMAIN 895 1000
FT DOMAIN 1001 1111
FT DOMAIN 1112 1211
FT DOMAIN 1218 1324
FT DOMAIN 1333 1436
FT DOMAIN 1437 1546
FT DOMAIN 1547 1649
FT DOMAIN 1650 1751
FT DOMAIN 1752 1855
FT DOMAIN 1856 1960
FT DOMAIN 1965 2068
FT DOMAIN 2069 2171
FT DOMAIN 2172 2277
FT DOMAIN 2278 2376
FT DOMAIN 2377 2482
FT DOMAIN 2483 2602
FT DOMAIN 2603 2706
FT DOMAIN 2707 2813
FT DOMAIN 2814 2933
FT CARBOHYD 217 217
FT CARBOHYD 256 256
FT CARBOHYD 402 402
FT CARBOHYD 584 584
FT CARBOHYD 1249 1249
FT CARBOHYD 1521 1521
FT CARBOHYD 1718 1718
FT CARBOHYD 1966 1966
FT CARBOHYD 2361 2361
FT CARBOHYD 2428 2428
FT CARBOHYD 2569 2569
FT CARBOHYD 2761 2761
FT CARBOHYD 2792 2792
FT CARBOHYD 2862 2862
SQ SEQUENCE 3298 AA; 346175 MW; 6E8BD28BE0795DB CRC64;

Query Match 0.5%; Score 15; DB 1; Length 3298;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 10  
MYC1\_CYPCA STANDARD; PRT; 394 AA.

```

AC Q90341;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYC I protein (C-MYC I).
GN CAM1.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_Taxid=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatopancreas;
RX MEDLINE=95180724; PubMed=7875594;
RA Zhang H., Okamoto N., Ikeda Y.;
RT "Two C-myc genes from a tetraploid fish, the common carp (Cyprinus
carpio).";
RL Gene 153:231-236(1995).
CC -1- FUNCTION: PARTICIPATES IN THE REGULATION OF GENE TRANSCRIPTION.
CC BINDS DNA BOTH IN A NON-SPECIFIC MANNER AND ALSO SPECIFICALLY TO
CC RECOGNIZES THE CORE SEQUENCE CAC(GA)TG. SEEMS TO ACTIVATE THE
CC TRANSCRIPTION OF GROWTH-RELATED GENES (BY SIMILARITY).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS AN HETERODIMER WITH MAX (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D37887; BAA07129.1; -.
DR HSSP; P25912; IHLO.
DR TRANSFAC; T03459; -.
DR InterPro; IPR001092; HLH basic.
DR InterPro; IPR003327; MYC-LZ.
DR InterPro; IPR002418; TF_MYC.
DR Pfam; PF00010; HLH; 1.
DR Pfam; PF02344; MYC-LZ; 1.
DR Pfam; PF01056; MYC N term; 1.
DR PRINTS; PR00044; DEUZIPRMYC.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; FALSE_NEG.
DR PROSITE; PS50888; HLH_2; 1.
KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW Glycoprotein; Multigene family.
FT DOMAIN 198 206
FT DOMAIN 207 224
FT DNA BIND 310 333
FT DOMAIN 324 363
FT DOMAIN 369 390
FT CARBOHYD 58 58
SQ SEQUENCE 394 AA; 44995 MW; 7146C81C920D71D0 CRC64;

Query Match 0.4%; Score 13; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 11  
AT32\_HUMAN STANDARD; PRT; 1211 AA.

AC 095450; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2) (Procollagen I/II  
 DE amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 DE I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2 OR PCNP OR PCNP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.  
 RC TISSUE=Skin;  
 RX MEDLINE=99347935; PubMed=10417273;  
 RA Collage A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,  
 RA Wierzelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,  
 RA Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.;  
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis  
 RT are caused by mutations in the procollagen I N-proteinase gene.";  
 RL Am. J. Hum. Genet. 65:308-317(1999).  
 CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
 CC to fibril assembly. Does not act on type III collagen. May also  
 CC play a role in development that is independent of its role in  
 CC collagen biosynthesis.  
 CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 CC alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
 CC at Ala-|-Gln.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
 CC COLLAGEN TYPE XIV (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=LPNPI;  
 CC IsoId=095450-1; Sequence=Displayed;  
 CC Name=SPNPI;  
 CC IsoId=095450-2; Sequence=VSP\_005497, VSP\_005498;  
 CC Note=Has no significant N-procollagen peptidase activity;  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON  
 CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.  
 CC -1- DOMAIN: THE SPACER DOMAIN WITH THE TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE (By  
 CC SIMILARITY).  
 CC -1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos  
 CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-  
 CC tissue disorder characterized clinically by severe skin fragility  
 CC and joint hypermobility and biochemically by the presence in skin  
 CC of procollagen incompletely processed at the N-terminus.  
 CC -1- SIMILARITY: Belongs to peptidase family M12B.  
 CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 4 TSP type-1 domains.  
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AJ003125; CA05880.1; -.  
 DR MEROPS: M12.301; -.  
 DR Genew; HGNC:218; ADAMTS2.  
 DR MIM; 604539; -.  
 DR MIM; 225410; -.  
 DR GO; GO:0008133; F:collagenase activity; TAS.

DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR002870; Pep\_M12B\_propep.  
 DR InterPro; IPR001590; ReprPolysin.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR000625; Zn\_M12B\_peptidase.  
 DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam; PF01421; ReprPolysin; 1.  
 DR Pfam; PF00090; Tsp\_1; 4.  
 DR SMART; SM00209; TSP1; 4.  
 DR PROSITE; PS50215; ADAM\_MEROPS; 1.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE; PS50214; DISINTEGRIN\_2; FALSE\_NEG.  
 DR PROSITE; PS50092; TSP1; 4.  
 DR PROSITE; PS00427; ZINC\_PROTEASE; FALSE\_NEG.  
 DR Hydroxylase; Metalloproteinase; zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;  
 KW Alternative splicing; Ehlers-Danlos syndrome.  
 FT SIGNAL 1 29  
 FT PROPEP 30 253  
 FT CHAIN 254 1211  
 FT DOMAIN 254 479  
 FT DOMAIN 480 560  
 FT DOMAIN 561 616  
 FT DOMAIN 618 722  
 FT SITE 651 693  
 FT DOMAIN 723 851  
 FT DOMAIN 854 912  
 FT DOMAIN 914 971  
 FT DOMAIN 975 1029  
 FT DOMAIN 1063 1090  
 FT DOMAIN 40 43  
 FT DOMAIN 185 188  
 FT METAL 408 408  
 FT ACT SITE 409 409  
 FT METAL 412 412  
 FT METAL 418 418  
 FT CARBOHYD 112 112  
 FT CARBOHYD 251 251  
 FT CARBOHYD 251 251  
 FT CARBOHYD 949 949  
 FT CARBOHYD 993 993  
 FT CARBOHYD 1031 1031  
 FT CARBOHYD 1098 1098  
 FT CARBOHYD 1145 1145  
 FT CARBOHYD 1150 1150  
 FT VASPLIC 544 566  
 FT FT  
 FT VASPLIC 567 1211  
 FT FT  
 FT SEQUENCE 1211 AA; 134722 MW; BECEPF25C3CAD2D CRC64;  
 SQ  
 Query Match 0.4%; Score 13; DB 1; Length 1211;  
 Best Local Similarity 100.0%; Pred. NO. 0.089;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 LLLLLLLLLLPPPL 30  
 DB 16 LLLLLLLLLLPPPL 28  
 RESULT 12  
 CD87 DROME STANDARD; FRT; 1967 AA.  
 AC Q9VG65; Q9VG65; 16-OCT-2001 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE Cadherin 87A precursor.  
 DE Cadherin 87A OR CG6977.  
 GN CAD87A OR CG6977.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyogota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OK NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Manamickandes P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wen K.H., Doyle C., Baxter E.G., Holt J., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busan S.D., Butler H., Cadieu E., Canter A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslø A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ilegam C.,  
 RA Jalali M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 RA Palazolo M., Piatan G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Styrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Waesman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=Berkley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.B.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [3]  
 RP SEQUENCE OF 435-1967 FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A *Drosophila* full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.  
 CC They preferentially interact with themselves in a homophilic  
 CC manner in connecting cells (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
 CC -1- SIMILARITY: Contains 14 cadherin domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).

CC EMBL; AE003693; AAF54717.3; -  
 CC EMBL; AY128505; AAM75098.1; -  
 CC HSSP; P5116; INCD.  
 DR Flybase: FBgn037963; Cad87A.  
 DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; NAS.  
 DR GO; GO:0016339; P:calcium-dependent cell-cell adhesion; NAS.  
 DR InterPro; IPR002126; Cadherin.  
 DR Pfam; PF00028; cadherin; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 14.  
 DR PROSITE; PS00232; CADHERIN 1; 3.  
 DR PROSITE; PS0268; CADHERIN 2; 14.  
 KM Hypothetical protein; Cell adhesion; Glycoprotein; Transmembrane;  
 KM Calcium; Calcium-binding; Repeat; Signal; Multigene family.  
 FT SIGNAL 1 ?  
 FT CHAIN ? 1967  
 FT DOMAIN ? 1767  
 FT TRANSMEM 1768 1788  
 FT DOMAIN 1789 1967  
 FT DOMAIN 20 124  
 FT DOMAIN 125 237  
 FT DOMAIN 238 350  
 FT DOMAIN 351 464  
 FT DOMAIN 465 661  
 FT DOMAIN 662 766  
 FT DOMAIN 767 870  
 FT DOMAIN 871 990  
 FT DOMAIN 991 1095  
 FT DOMAIN 1096 1203  
 FT DOMAIN 1204 1310  
 FT DOMAIN 1311 1423  
 FT DOMAIN 1424 1545  
 FT DOMAIN 1546 1669  
 FT CARBOHYD 31 31  
 FT CARBOHYD 69 69  
 FT CARBOHYD 195 195  
 FT CARBOHYD 416 416  
 FT CARBOHYD 722 722  
 FT CARBOHYD 753 753  
 FT CARBOHYD 1031 1031  
 FT CARBOHYD 1041 1041  
 FT CARBOHYD 1103 1103  
 FT CARBOHYD 1155 1155  
 FT CARBOHYD 1209 1209  
 FT CARBOHYD 1317 1317  
 FT CARBOHYD 1341 1341  
 FT CARBOHYD 1484 1484  
 FT CARBOHYD 1568 1568  
 FT CARBOHYD 1683 1683  
 FT CARBOHYD 763 763  
 FT CONFLICT 1102 1102  
 FT CONFLICT 1642 1642  
 FT CONFLICT 1682 1682  
 SQ SEQUENCE 1967 AA; 216689 MW; 7164EB5E80CF41 CRC64;  
 Query Match 0.4%; Score 13; DB 1; Length 1967;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 TVLDVNDNPVPE 922  
 |||  
 DB 1534 TVLDVNDNPVPE 1546  
 |||  
 RESULT 13

```

RLA0 METKA
ID METKA STANDARD; PRT; 357 AA.
AC 08TX50;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acidic ribosomal protein P0 homolog (L10B).
GN RPLP0 OR MK0826.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
NCBI_Taxid=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RA MEDLINE=21927647; PubMed=11930014;
RA Sleasore A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcharbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natchale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.,
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
CC -1- FUNCTION: Ribosomal protein P0 is the functional equivalent of
CC E. coli protein L10.
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; AEO10373; AAM02039.1; ALT_INIT.
DR HAMAP; MF_00280; -1.
DR InterPro; IPR001790; Ribosomal_L10.
DR Pfam; PF00466; Ribosomal_L10; 1.
DR K01001; Ribosomal protein, Complete proteome.
SQ SEQUENCE 357 AA; 39250 MW; 470294320ADBBESC CRC64;

Query Match 0.4%; Score 12; DB 1; Length 357;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 EEEEEEEEA 2754
DB 338 EEEEEEEEA 349

RESULT 14
PTPX MACNE STANDARD; PRT; 1013 AA.
ID PTPX MACNE
AC 002655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Protein-tyrosine phosphatase X precursor (EC 3.1.3.48) (R-PTP-X)
DE (M1851).
GN PTPRX2.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
NCBI_Taxid=9545;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RA MEDLINE=97254813; PubMed=9100223;
RA Lagasee J., Jellinek L., Sexson S., Lofton-Day C.B., Breining J.,
RA Sheppard P., Kindesvogel W., Hagopian W.A.,
RT "An islet-cell protein tyrosine phosphatase is a likely precursor to

```

```

RT the 37-kDa autoantigen in type 1 diabetes: human and macaque
RT sequences, tissue distribution, unique and shared epitopes, and
RT predictive autoantibodies."
RL Mol. Med. 3:163-173 (1997).
CC -1- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND
CC PANCREATIC ENDOCRINE CELLS.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: BRAIN, PROSTATE, PANCREATIC ISLETS. LOWER
CC EXPRESSION IN SPINAL CORD, THYROID, ADRENAL MEDULLA AND
CC GASTROINTESTINAL TRACT. MULTIPLE PROTEOLYTIC CLEAVAGE AT
CC -1- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT
CC CONSECUTIVE BASIC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; U91574; AAC51186.1; -.
DR HSSP; P18052; TYPO.
DR InterPro; IPR000387; TYR phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTPHPTASR.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
DR K01001; Hydroxylase; Receptor; Glycoprotein; Signal; Transmembrane.
FT SIGNAL 1 19
FT CHAIN 20 1013
FT DOMAIN 20 613
FT TRANSMEM 614 634
FT DOMAIN 635 1013
FT DOMAIN 768 1002
FT ACT SITE 943 943
FT SITE 425 426
FT CARBOHYD 562 562
SQ SEQUENCE 1013 AA; 11190 MW; 4808D43937A2EF59 CRC64;

Query Match 0.4%; Score 12; DB 1; Length 1013;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 PLLLLLLLLPP 27
DB 4 PLLLLLLLLPP 15

RESULT 15
PTPX HUMAN STANDARD; PRT; 1015 AA.
ID PTPX HUMAN
AC 092932; Q92662;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-tyrosine phosphatase X precursor (EC 3.1.3.48) (R-PTP-X)
DE (Islet cell autoantigen related protein) (ICAPAR) (IAR) (Phogin).
GN PTPRX2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA MEDLINE=97032784; PubMed=8878534;

```

RA Kawasaki E., Hutton J.C., Eisenbarth G.S.;  
 RT "Molecular cloning and characterization of the human transmembrane  
 RT protein tyrosine phosphatase homologue, plogrin, an autoantigen of  
 RT type 1 diabetes.";  
 RL Biochem. Biophys. Res. Commun. 227:440-447(1996).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97127415; PubMed=8954911;  
 RA Smith P.D., Barker K.T., Wang J., Lu Y.-J., Shipley J., Crompton M.R.;  
 RT "ICAR, a novel member of a new family of transmembrane, tyrosine  
 RT phosphatase-like proteins.";  
 RL Biochem. Biophys. Res. Commun. 229:402-411(1996).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Brain, and Pancreas;  
 RX MEDLINE=96394649; PubMed=8798755;  
 RA Cui L., Yu W.-P., de Aizpurua H.J., Schmidt R.S., Pallen C.J.;  
 RT "Cloning and characterization of islet cell antigen-related protein-  
 RT tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen  
 RT in insulin-dependent diabetes.";  
 RL J. Biol. Chem. 271:24817-24823(1996).  
 CC -1- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND  
 CC PANCREATIC ENDOCRINE CELLS. OPTIMUM ACTIVITY IS MEASURED AT PH  
 CC 4.5.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND PANCREAS. LOWER  
 CC LEVELS IN TRACHEA, PROSTATE, STOMACH AND SPINAL CHORD.  
 CC -1- DOMAIN: THE CYTOPLASMIC DOMAIN APPEARS TO CONTRAIN THE  
 CC AUTOANTIGENIC EPITOPES.  
 CC -1- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT  
 CC CONSECUTIVE BASIC RESIDUES.  
 CC -1- DISEASE: AUTOANTIGEN IN INSULIN-DEPENDENT DIABETES MELLITUS  
 CC (IDDM).  
 CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U66702; AAC50742.1; -;  
 DR EMBL: Y08569; CAAG9880.1; -;  
 DR EMBL: AF007555; AAB63600.1; -;  
 DR PIR: JC5062; JC5062.  
 DR PIR: JC5263; JC5263.  
 DR HSP: P18052; IYFO.  
 DR GeneW: HGNC:9677; PTPRN2.  
 DR MIM: 601698; -;  
 DR GO: GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO: GO:0005001; F:transmembrane receptor protein tyrosine pho. . .; TAS.  
 DR GO: GO:0006470; P:protein amino acid dephosphorylation; TAS.  
 DR InterPro: IPR000387; TYR phosphatase.  
 DR InterPro: IPR000242; TYR\_PP.  
 DR Pfam: PF00102; Y\_phosphatase; 1.  
 DR PRINTS: PR00700; PRTYPHPTASE.  
 DR SMART: SM00194; PTPC; 1.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR HydroLase; Receptor; Glycoprotein; Signal; Transmembrane.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1015 PROTEIN-TYROSINE PHOSPHATASE X.  
 FT DOMAIN 22 615 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 616 636 POTENTIAL.  
 FT DOMAIN 637 1015 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 770 1004 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 945 945

FT	SITE	427	428	CLEAVAGE SITE (POTENTIAL).
FT	CARBOHYD	564	564	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MUTAGEN	945	945	C->S: LOSS OF ACTIVITY.
FT	CONFLICT	208	208	P -> S (IN REF. 2 AND 3).
FT	CONFLICT	247	247	S -> G (IN REF. 2).
FT	CONFLICT	323	323	G -> R (IN REF. 2).
FT	CONFLICT	325	325	S -> N (IN REF. 2).
SO	SEQUENCE	1015 AA;	111281 MW;	82AC3CC9F59AC97 CRC64;

Query Match 0.4%; Score 12; DB 1; Length 1015;  
 Best Local Similarity 100.0%; Pred. No. 0.45;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	16 PLALLLLLLPP 27
Db	6 PLALLLLLLPP 17

Search completed: February 11, 2004, 16:06:43  
 Job time : 28 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:03:00 ; Search time 75 Seconds  
(without alignments)  
10057.170 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 2923  
Sequence: 1 MESPATGVLPPTPPPLLL.....AGTVDESSGSRFLPFLH 2923

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organella:\*  
9: SP\_phase:\*  
10: SP\_plant:\*  
11: SP\_rodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriaph:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	1.1	32	11	Q9J1U2
2	24	0.8	25	11	Q9J1U4
3	19	0.7	34	6	O46539
4	18	0.6	138	13	Q9YQ08
5	18	0.6	1766	13	Q8AM45
6	15	0.5	138	13	Q9YQ09
7	13	0.4	311	11	Q8VE43
8	13	0.4	382	11	Q9D0M2
9	13	0.4	382	11	Q8C1A0
10	13	0.4	414	6	Q8HZR0
11	13	0.4	436	11	Q8BNR6
12	13	0.4	467	4	Q9BTY2
13	13	0.4	467	4	Q8BNK4
14	13	0.4	467	4	Q9JUM5
15	13	0.4	633	6	Q8HZR1
16	12	0.4	176	11	Q8QZT0

17	12	0.4	176	11	Q8C1M2	Q8C1M2 mus musculus
18	12	0.4	239	6	Q9XSK7	Q9XSK7 bos taurus
19	12	0.4	271	11	Q8CDV2	Q8CDV2 mus musculus
20	12	0.4	293	10	Q9R1U5	Q9R1U5 arabidopsis
21	12	0.4	315	4	Q9EHG7	Q9EHG7 homo sapien
22	12	0.4	330	11	Q8C1A4	Q8C1A4 mus musculus
23	12	0.4	334	11	Q8C1U8	Q8C1U8 mus musculus
24	12	0.4	401	11	Q9CWC8	Q9CWC8 mus musculus
25	12	0.4	401	11	Q8C1U9	Q8C1U9 mus musculus
26	12	0.4	401	11	Q8BSN5	Q8BSN5 mus musculus
27	12	0.4	403	4	Q9BUL5	Q9BUL5 homo sapien
28	12	0.4	538	4	Q9NM07	Q9NM07 homo sapien
29	12	0.4	540	4	Q8N0Z9	Q8N0Z9 homo sapien
30	12	0.4	606	6	Q97554	Q97554 oryctolagus
31	12	0.4	608	10	Q9SU12	Q9SU12 arabidopsis
32	12	0.4	629	11	Q9RT51	Q9RT51 mus musculus
33	12	0.4	793	11	Q70246	Q70246 mus musculus
34	12	0.4	813	11	Q8B0C3	Q8B0C3 mus musculus
35	12	0.4	896	11	Q9R093	Q9R093 ratius norv
36	12	0.4	900	17	Q8PUB5	Q8PUB5 methanosarc
37	12	0.4	986	4	Q9Y419	Q9Y419 homo sapien
38	12	0.4	986	4	Q8N4T5	Q8N4T5 homo sapien
39	12	0.4	998	4	Q9Y416	Q9Y416 homo sapien
40	12	0.4	1032	11	Q9Z1T4	Q9Z1T4 ratius norv
41	12	0.4	1270	4	Q96JH2	Q96JH2 homo sapien
42	12	0.4	1378	11	Q35850	Q35850 ratius norv
43	12	0.4	4169	4	Q8CTU4	Q8CTU4 homo sapien
44	11	0.4	59	11	Q9R1N5	Q9R1N5 mus musculus
45	11	0.4	75	10	Q9ST05	Q9ST05 arabidopsis

## ALIGNMENTS

### RESULT 1

ID	Q9J1U2	PRELIMINARY;	PRT;	32 AA.
AC	Q9J1U2;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Flamingol (Fragment).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Fischer344; TISSUE=Testis;			
RX	MEDLINE=20114371; PubMed=10650949;			
RA	Johnson K.J., Patel S.R., Boekelheide K.;			
RT	"Multiple cadherin superfamily members with unique expression profiles are produced in rat testis";			
RL	Endocrinology 141:675-683(2000).			
DR	EMBL; AF177697; AAF87072.1; -.			
FT	NON TER			
FT	NON TER			
SQ	SEQUENCE 32 AA; 3554 MW; 60ABB46A7B62616B CRC64;			
Query Match	1.1%; Score 32; DB 11; Length 32;			
Best Local Similarity	100.0%; Pred. No. 3,1e-23;			
Matches	32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	777 DOVSYTLATFARDNGIPKSDTTTLRLTVNDV 808			
DB	1 DOVSYTLATFARDNGIPKSDTTTLRLTVNDV 32			
RESULT 2				
Q9J1U4	PRELIMINARY;			
ID	Q9J1U4	PRT;	25 AA.	
AC	Q9J1U4;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			



DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE Flamingol (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Fischer344; TISSUE=Testis;  
 RX MEDLINE=20114371; PubMed=10650949;  
 RA Johnson K.J., Patel S.R., Boekelheide K.;  
 RT "Multiple cadherin superfamily members with unique expression profiles  
 are produced in rat testis.";  
 RL Endocrinology 141:675-683(2000).  
 DR EMBL; AF177695; AAF87070.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 25 25  
 SQ SEQUENCE 25 AA; 2605 MW; 09940DE0F9P95794 CRC64;

Query Match 0.8%; Score 24; DB 11; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-15;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 RAODGGRPPISNVSGLYTVQYLDI 498  
 DB 2 RAODGGRPPISNVSGLYTVQYLDI 25

## RESULT 3

046539 PRELIMINARY; PRT; 34 AA.  
 AC 046539;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
 DE Cadherin-like protein (Fragment).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 NCBI\_TaxID=9823;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RA Lutz D.A., Zheng J.J.;  
 RT "Expression of multiple cadherins in adult retinal pigment epithelial  
 (RPE) cells.";  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF037464; AAB92582.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 34 34  
 SQ SEQUENCE 34 AA; 3648 MW; 5FCF1719B718F855 CRC64;

Query Match 0.7%; Score 19; DB 6; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 781 YTLAITARDNGIPKSDPT 799  
 DB 5 YTLAITARDNGIPKSDPT 23

## RESULT 4

09YGO8 PRELIMINARY; PRT; 138 AA.  
 AC 09YGO8;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Cadherin (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.  
 NCBI\_TaxID=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Miskevich F., Zhu Y., Ranscht B., Sanes J.R.;  
 RT "Expression of Multiple Cadherins and Catenins in the Chick Optic  
 Tectum.";  
 RL Mol. Cell. Neurosci. 0:0-0(1999).  
 DR EMBL; AF098479; AAD16269.1; -.  
 DR HSSP; P09803; 1EDH.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 1.  
 DR SMART; SM00112; Ca; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 1.  
 DR PROSITE; PS0268; CADHERIN\_2; 2.  
 KM Calcium; Calcium-binding; Cell adhesion; Glycoprotein.  
 FT NON\_TER 1 1  
 FT NON\_TER 138 138  
 SQ SEQUENCE 138 AA; 14523 MW; E03B8D50074F48E6 CRC64;

Query Match 0.6%; Score 18; DB 13; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 NDNAPIFVSTPPOATYLE 516  
 DB 34 NDNAPIFVSTPPOATYLE 51

## RESULT 5

08AW45 PRELIMINARY; PRT; 1766 AA.  
 AC 08AW45;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Sl:b6L08.1 (Novel protein similar to cadherin, EGF LAG seven-pass  
 G-type receptor (CELSR)) (Fragment).  
 GN Sl:B6L08.1.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Brachydanio rerio (Zebrafish); Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_TaxID=7955;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Babbage A.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL645790; CAD58738.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 1766 1766  
 SQ SEQUENCE 1766 AA; 197485 MW; 80CAAAA18F5A497F CRC64;

Query Match 0.6%; Score 18; DB 13; Length 1766;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 VLPPDNDICLRPCENYM 1241  
 DB 117 VLPPDNDICLRPCENYM 134

## RESULT 6

09YGO9 PRELIMINARY; PRT; 138 AA.  
 AC 09YGO9;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Cadherin (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Miskelich F., Zhu Y., Ranscht B., Sanes J.R.;
RT "Expression of Multiple Cadherins and Catenins in the Chick Optic
RL Tectum.";
DR EMBL: AF098478; AAD1268.1; -.
DR HSRP; P15116; INCI.
DR InterPro; IPR002126; Cadherin.
DR SMART; SM00028; cadherin; 1.
DR PROSITE; PS00232; CADHERIN_1; 1.
DR PROSITE; PS00268; CADHERIN_2; 2.
KM Calcium-binding; Cell adhesion; Glycoprotein.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15282 MW; A68C982E27CBB290 CRC64;

Query Match 0.5%; Score 15; DB 13; Length 138;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 PPFTSVLOISATPRD 843
DB 53 PPFTSVLOISATPRD 67

RESULT 7
O8VE43 PRELIMINARY; PRT; 311 AA.
AC O8VE43;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Hypothetical protein.
GN BC019776.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RT Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; BC019776; AAH19776.1; -.
DR EMBL; BC024445; AAH24445.1; -.
DR EMBL; BC026646; AAH26646.1; -.
DR EMBL; AK034855; BAC28856.1; -.
DR MGD; MGI:2384806; BC019776.
KM Hypothetical protein.
SQ SEQUENCE 311 AA; 34529 MW; C806268E2B6DFA0 CRC64;

Query Match 0.4%; Score 13; DB 11; Length 311;

```

```

Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 PPPPLLLLLL 25
DB 28 PPPPLLLLLL 40

RESULT 8
O9DM2 PRELIMINARY; PRT; 382 AA.
ID O9DM2;
AC O9DM2;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE 2310021G01RIK protein (Similar to C-MYC target JF01).
GN 2310021G01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukushima Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK011289; BAB27519.1; -.
DR EMBL; AK075882; BAC36027.1; -.
DR MGD; MGI:1914203; 2310021G01RIK.
SQ SEQUENCE 382 AA; 43837 MW; 5192063588CD40 CRC64;

Query Match 0.4%; Score 13; DB 11; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2740 SDSSEEEEEEE 2752
DB 103 SDSSEEEEEEE 115

RESULT 9
O8CIA0 PRELIMINARY; PRT; 382 AA.
ID O8CIA0;
AC O8CIA0;
DT 01-MAR-2003 (TRENBLREL. 23, Created)

```



RT "HRI human cDNA sequencing project";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A075458; BAC11633.1; -  
 DR InterPro; IPR000933; Alpha\_L\_fucos.  
 DR Pfam; PF01120; Alpha\_L\_fucos; 1.  
 DR PRINTS; PR00741; GLHYDRLASE2.  
 DR PROSITE; PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 KW Hypothetical protein: 54044 MW; 933BDB9858BDD5 CRC64;  
 SQ SEQUENCE 467 AA; 54044 MW; 933BDB9858BDD5 CRC64;  
 Query Match 0.4%; Score 13; DB 4; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 12 PLLLLLLLLPPP 24

RESULT 14  
 Q9UUM5 PRELIMINARY; PRT; 467 AA.  
 ID Q9UUM5  
 AC Q9UUM5; 01-MAY-2000 (TREMELREL. 13, Created)  
 DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMELREL. 23, Last annotation update)  
 DE DJ20N2.5 (Novel protein similar to fucosidase, alpha-L-1, tissue  
 (EC 3.2.1.51, alpha-L-fucosidase fucosylolase)).  
 GN DJ20N2.5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kay M.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031320; CMB53746.1; -  
 DR InterPro; IPR000933; Alpha\_L\_fucos.  
 DR Pfam; PF01120; Alpha\_L\_fucos; 1.  
 DR PRINTS; PR00741; GLHYDRLASE2.  
 DR PROSITE; PS00385; ALPHA\_L\_FUCOSIDASE; 1.  
 SQ SEQUENCE 467 AA; 54066 MW; A42AA6B369A3AC39 CRC64;  
 Query Match 0.4%; Score 13; DB 4; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 12 PLLLLLLLLPPP 24

RESULT 15  
 O8HZRI PRELIMINARY; PRT; 633 AA.  
 ID O8HZRI  
 AC O8HZRI; 01-MAR-2003 (TREMELREL. 23, Created)  
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)  
 DE Cyclooxygenase.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chandrasekharan N.V., Dai H., Roos K.L.T., Evanson N.K., Tomsik J.,  
 Elton T.S., Simmons D.L.;  
 RT "COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and  
 other analgesic/antipyretic drugs: Cloning, structure, and  
 expression.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:13926-13931(2002).  
 DR EMBL; AF535138; AAN33049.1; -  
 SQ SEQUENCE 633 AA; 72502 MW; D338221B976BDB7E CRC64;  
 Query Match 0.4%; Score 13; DB 6; Length 633;  
 Best Local Similarity 100.0%; Pred. No. 0.0036;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 PLLLLLLLLPPP 28  
 DB 41 PLLLLLLLLPPP 53

Search completed: February 11, 2004, 16:08:09  
 Job time : 76 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 05:45:30 ; Search time 22141 Seconds  
(without alignments)  
5400.788 Million cell updates/sec

Title: US-09-916-849a-3  
Perfect score: 2923  
Sequence: 1 MRSPTGVLPPLPPPLLL.....ACTVDESSGSSEPLFFNPLH 2923

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:  
MODEL=frame+ p2n.model -DEV=xlh  
O=/cgn2.1/USPTO.spool/US09916849/runat.11022004.143917.21352/app.query.fasta.1.3079  
-DB=GenEmbl -QFMT=fastap -SUPPLX=011go.rge -MINMATCH=0.1 -IOOPT=0 -IOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=011go -TRANS=human0.cdi -LIST=45  
-DOCLIGN=200 -THR.SCOR=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro  
-NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849@cgn.1.15745 @runat.11022004.143917.21352 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEBQUERY -NEG.SCOR=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_scs:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mmu:\*  
20: em\_mu:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_to:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ey:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2923	100.0	8772	6	AX224128 Sequence
2	2923	100.0	10531	6	AX549238 Sequence
3	2923	100.0	10531	9	AF234887 Homo sapi
4	2837	97.1	8871	6	AX224126 Sequence
5	2408	82.4	8924	9	DB7469 Human mRNA
6	1958	67.0	9321	6	AX367102 Sequence
7	1103	37.7	24370	6	AX646739 Sequence
8	1103	37.7	24370	9	AB065955 Homo sapi
9	1103	37.7	169241	9	AL390252 Homo DNA
10	1103	37.7	227194	2	BX284647 Homo sapi
11	1091	37.3	116603	2	AL157901 Homo sapi
12	609	20.8	2821	9	AK091437 Homo sapi
13	482	16.5	89129	9	AL672200 Mouse DNA
14	482	16.5	234932	2	AC093365 Mus muscu
15	347	11.9	3912	6	AX098823 Sequence
16	305	10.4	235320	2	AC106175 Rattus no
17	305	10.4	235327	2	AC095463 Rattus no
18	305	10.4	248059	2	AC113756 Rattus no
19	305	10.4	254975	2	AC095125 Rattus no
20	209	7.2	932	9	HSA23766 Homo sapi
21	208	7.1	1007	10	AY121289 Homo sapi
22	187	6.4	6699	10	AB011529 Rattus no
23	168	5.7	8950	10	AB028499 Mus muscu
24	116	4.0	2653	10	AF031573 Mus muscu
25	109	3.7	253371	2	AC143681 Macaca mu
26	92	3.1	285	6	AX524129 Sequence
27	92	3.1	285	6	AX552867 Sequence
28	87	3.0	3343	10	BC005499 Mus muscu
29	81	2.8	271	11	G16013 human STS C
30	53	1.8	300	11	G37263 SHGC-57403
31	44	1.5	169860	2	AC110425 Rattus no
32	43	1.5	266451	2	AC143184 Macaca mu
33	34	1.2	132775	10	AL671899 Mouse DNA
34	32	1.1	97	10	AF177697 Rattus no
35	29	1.0	167579	5	AL645789 zebrafish
36	24	0.8	77	10	AF177695 Rattus no
37	23	0.8	10195	6	AF0094 Sequence 4
38	23	0.8	10919	10	AF031572 Mus muscu
39	23	0.8	238919	2	AC139513 Rattus no
40	23	0.8	253297	2	AC134076 Rattus no
41	23	0.8	260331	2	AC135400 Rattus no
42	23	0.8	263730	2	AC094314 Rattus no
43	21	0.7	135	6	AR025222 Sequence
44	21	0.7	135	6	AR070009 Sequence
45	21	0.7	135	6	IS1001 Sequence 15

RESULT 1

## ALIGNMENTS

AX224128  
LOCUS AX224128 8772 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 3 from Patent WO0161003.  
ACCESSION AX224128  
VERSION AX224128.1 GI:15554400  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1  
AUTHORS Testa, T.T.  
TITLE The human homologue of flamingo  
JOURNAL Patent: WO 0161003-A 3 23-AUG-2001;  
SMITHKLINE BEECHAM PLC (GB)  
FEATURES  
source 1. 8772  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 1696 a 2814 c 2564 g 1698 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 8772  
Score: 2923.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-916-849a-3 (1-2923) x AX224128 (1-8772)

Qy 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
Db 1 ATGGGAGGCCCGGCGACCGGCTCCCTCCCAAGCGCGCGCGCTGCTGCTG 60  
Qy 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
Db 61 TTGCTGCTGCTGCTGCGCGCGCACTATTGGAGACCAAGTGGGGCCCTGCTCTTG 120  
Qy 41 G1ySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTyrPleuGlyProSer 60  
Db 121 GGGTCCAGGGAGCAGGCTCTTCGGGGGCTGCGCGCCCAAGGAGCTGCTGCAATCC 180  
Qy 61 SerAlaSerAsnLeuTyrPleuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80  
Db 181 TCGAGCTCGAAGCTCTGGCTTACCAAGCGGCTGAGGGATGGGGCACTGAGCTGACT 240  
Qy 81 G1yHisLeuValProHisAspGlyLeuArgValTyrCysProGlnSerGlnHis 100  
Db 241 GGGCACTCGTATCCCAAGCAGATGGCTGAGGGTTCGATGTCAGAAATCCAGGGCCAT 300  
Qy 101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyTyrLeu 120  
Db 301 ATTCCCTCAACCAAGCTCTCGAAGGCTGCGCTGAGCTGCTGCTGCGGATTTGGA 360  
Qy 121 G1yHisLeuSerProGlnGlyValaLeuThrLeuProGlnGlnHisProCysLeuVala 140  
Db 361 GGGCACTTCCTCCCAAGGGGCAAGCTCACTGCGGAGAGGACCCGCTGTTAAAGCT 420  
Qy 141 ProArgLeuArgCysGlnSerCysAlaLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 421 CCAAGGCTCAGATCCAGTCTCTGCAAGCTGCAAGGCCCCCGGCTCAGGGGAGAA 480  
Qy 161 ArgSerProGlnGlnSerLeuGlyValaArgArgLeuAsnValaAsnThrAlaProGln 180  
Db 481 AGGTCAACCAAGAGTCTCTGGTGGGCTGAGAAAGAAATGAATACAGCCCCCAG 540  
Qy 181 PheGlnProProSerTyrGlnAlaThrValProGlnGlnGlnProAlaGlyThrProVal 200  
Db 541 TTCAGGCCCCCACTACCAAGGCCAAGTGCAGGAAACAGGAGGAGGACCCCTGTT 600

Qy 201 AlaSerLeuArgAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlnTyrThrMet 220  
Db 601 GCATCCCTGAGGGCCATCGACCCGAGCAGGGGTGAGGAGCTGACTGAGTACACCAAG 660  
Qy 221 AspAlaLeuPheAspSerArgSerArgSerGlnPhePheSerLeuAspProValThrGlyAla 240  
Db 661 GATGCCCTCTTGATATGCCGCTCCCAACAGTCTTCTCTGAGACCAAGTCACTGGTGA 720  
Qy 241 ValThrThrAlaGlnGlnLeuAspArgGlnTyrThrLysSerThrHisValPheArgValThr 260  
Db 721 GTTACCAACGCGAGAGAGCTGATCTGAGACCAAGACCAACCACTCTTCAGGGTCAAG 780  
Qy 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
Db 781 GGCACAGACCAAGGCGATGCCCGACCAAGATGCCCTGGCTACACTCACTCTGTTACT 840  
Qy 281 AspThrAspAspHisAspProValPheGlnGlnGlnGlnTyrThrLysGlnSerLeuArgGlu 300  
Db 841 GACACCAATGACCATGACCTGTTGTCAGACAGAGAGATGACAGAGAGCTTCAGGGAG 900  
Qy 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 901 AACCTGAGGTTGGCTATGAGGTGCTCACTGTCAAGGGCACCGATGTATGCCCTCC 960  
Qy 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340  
Db 961 AATGCCAATATTCGTATCCGCTGCTGAGGGGCTGCGGGGAGGCCCTCTGAAGTCTTT 1020  
Qy 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnVal 360  
Db 1021 GAGATGACCTGCTGCTGGGGTATCCAGACCCGTTGGCTTGATTCGGAGAGGGTG 1080  
Qy 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
Db 1081 GATCTTACACGTACGCTAGAGGAGCAATGACAGAGGTGCGGAGCCGGGCTCTCGAGAT 1140  
Qy 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAspAspAlaProGlnPheSer 400  
Db 1141 ACCACAGCCGCTGTTTCTCTTCTGAGAGATGACAAATGAATATGCCCCCAAGTTTACT 1200  
Qy 401 GlnValArgTyrValValGlnValArgGlnAspValThrProGlyValaProValLeuArg 420  
Db 1201 GAGAGCGCTATGTGTCCAGGTGAGGAGATGTATCTCCAGGGGCCCACTATCTCCA 1260  
Qy 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValaHisTyrSerIleMetSer 440  
Db 1261 GTCAAGCCTCGGATCGAAGCAAGGGGAGCAATGCCGTGTGCACTATAGCATGAGAT 1320  
Qy 441 G1yAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAspValaSer 460  
Db 1321 GGCATGCTCGGGGACAGTTTATCTGATGTGCCAGACTGAGCTCTGATGTGTGAGC 1380  
Qy 461 ProLeuAspTyrGlnThrThrLysGlnTyrThrLeuArgValArgAlaGlnAspGly 480  
Db 1381 CCTCTTGACTATGAGCAGACCAAGAGTACACCTTCAGGGTGCAGACAGATGTGGC 1440  
Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1441 CGTCCCACTCTCTATGTCTCTGGCTTGAGACAGTACAGGTCTGTGATATCAACGAC 1500  
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValaProLeu 520  
Db 1501 AATGCCCATATCTTCCTGACAGCCCTTTCAGGGCTACTGTCTCTGAGAGCGTCCCTTA 1560  
Qy 521 G1yTyrLeuValLeuHisValaGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1561 GGCTTACTGTCTTCATGTCCAGGCTATCGAGCTGATCGATGCTGTGATCAATGCCGCTG 1620  
Qy 541 GlnTyrArgLeuAlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly 560  
Db 1621 GAATACCGCTGCTGCTGGGAGCAATGACTTCCCTTCAACATCAATGAGCAAGGC 1680  
Qy 561 TrpIleSerValaIleAlaGlnLeuAspArgGlnGlnValaAspPheTyrSerPheGlyVal 580

Db	1681	TGAACTCTGTCGGCTGCTGAATCGACCGGGAGAAATTGATTCTACCTTTGGGTA	1740
Qy	561	GIWIAAAGAAAPHISGLYTHrProAlaLeuThraIAserAlaSerValIthrVal	600
Db	1741	GAAGCTCAGACCAAGGACCTCCAGACCTCACTGCTCGGCGCAAGTCAAGGTGACTGTC	1800
Qy	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyThrValArgLeuAsn	620
Db	1801	CTGAGATTCACAGCAACAATCAATCACTTTACCAACAGATACAGTGGGCTCAAT	1860
Qy	621	GIuAspAlaAlaValGlyThrSerValIthrValSerAlaValAspArgAspAlaHis	640
Db	1861	GAGATGACGCTGTGGGACCAAGCTGGTGAAGGTGACGCTGTGACGATGATCTCAT	1920
Qy	641	SerValIleThrTyrglnIleThrseryIAsnThrArgAsnArgPheSerIleThrser	660
Db	1921	AGTGTCAATCACTTACAGATCAACAAGTGCATATCTCAAAACCGCTTCCATACCAACG	1980
Qy	661	GlnSergIyGIyGIyLeuValSerLeuAlaLeuProLeuAspTyTyLeuGIuArgGln	680
Db	1981	CAAGGTGGTGGGTGGCTGTATCCCTTGCCCTGCACCTAGACTACAACTTGAAGGGAG	2040
Qy	681	TyrValIleuAlaValThraIAserAspGIyThraGlnAspThraIleGlnIleVal	700
Db	2041	TATGTGTGGCTGTACCGCTCCCATGTGACCTCGGCGAGCACGGCACAACTTGTGGTG	2100
Qy	701	AsnValThraAspAlaAsnThrHisArgProValPheGlnAsrSerHisTyThrValAsn	720
Db	2101	AATGTACCGACGCGCAACCCATGTCTCTGTCTTCAAGACTCCCATTAATACGTGAT	2160
Qy	721	ValAsnGIuAspArgProAlaGIyThrThraValIleuIleSerAlaThraAspGIuAsp	740
Db	2161	GTTAATAGGAGACCGGCGCGGAGGACCAACGGTGGTGTATCAGGCGCCACGATGAGAC	2220
Qy	741	ThrGIyGIuAsnAlaArgIleThrTyrrPheMetGIuAspSerIleProGlnPheArgIle	760
Db	2221	ACGGGTGGAATGCCCGCATCACTCACTCAATGAGGACACATCTCCCAATTCGGCATC	2280
Qy	761	AspAlaAspThrGIyAlaValThrThrGlnAlaGluLeuAspTyGIuAspGlnValSer	780
Db	2281	GATGACAGCAAGGGGGCTGTCAACCAACGAGGTGAGTGACTACGAAGACCAAGTGTCT	2340
Qy	781	TyrThrIleuAlaIleThraIAsrAspAsnGIyIleProGlnIlySerAspThrThrTy	800
Db	2341	TACACCTCGGCATTACTGTCTCGGACATGGCATTTCCCAAGATCGACACCACTTAC	2400
Qy	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTy	820
Db	2401	CTGAGAGATCCTGGTGAACGACGTGAATACATAGCCCTCACTTCTCGAGACTCTTAC	2460
Qy	821	GlnGlySerValTyrgIuAspValProProPheThrserValLeuGlnIleSerAlaThr	840
Db	2461	CAGGGCAGATGTCATGAGAGATGTGCACTTCCTACAGCGTCCGAGATCTCAGCACT	2520
Qy	841	AspArgAspSergIyLeuAsnGIyAArgValPheTyrrThrPheGlnGlyIyAspAspGIy	860
Db	2521	GATGTGTATCTTGAATTAAATGCAAGGAGTCTTCAACCTTCAAGGAGGCGAAGATGGA	2580
Qy	861	AspGIyAspPheIleValGIuSerThrseryIleValArgThrLeuArgArgLeuAsp	880
Db	2581	GACGGTACTTTATATTGTGAGTCCACGTCAGGCAATCGTGCAAGGCTACGAGGCTGGAT	2640
Qy	881	ArgGIuAsnValAlaGlnTyThrValIleuArgAlaTyraIaValaAspIyGlyMetProPro	900
Db	2641	CGAGGAACCGTGGCCAGATATGTCTTGGGGCATATGACAGTGCACAAAGGGGATCCCCCA	2700
Qy	901	AlaArgThrProMetGluValThrValIthrValIleuAspValAsnAspAsnProProVal	920
Db	2701	GCCGCGCACACTATGGAAGTAGCACTGTGTGTTGATGTGATGACAAATCCCCCTGTCTC	2760
Qy	921	PheGluGlnAspGIuPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal	940

Dp	2761	TTTGAAGCAGGATGAGTTGATGTTGTTTGTGGAAGAAGAACAGCCCATTTGGGCTAAGCCGTG	2820
Qy	941	AlAaTgValThrAlaThrAaPProAaPgiUgiYThraAnaIaGInIIeMeTyrGInIle	960
Dp	2821	GCCCCGGGTCAACAGCCACTGACCCCGGAGGAAGGCCAACATGCCAGATTATATGACCAATT	2880
Qy	961	ValGInGIaAnaIIeProGIuValPheGIaLeuAaPILePheSaRGIuLeuThraAla	980
Dp	2881	GTGAAGGGCAACATCCCTGAAGTCTTTCAGCTGACATCTTCTCCGGGAAGCTACAGCC	2940
Qy	981	LeuValAaPLeuAaPTRYGIuAaPaRgProGIuTYrValLeuValIIeGInIaIaThrSer	1000
Dp	2941	CTGGTAAAGCTTAACTACAGAGAACCGGCGCTGAAGTACGTCCTGGTCAACAGGCCATGCA	3000
Qy	1001	AlAaProLeuValSerAaRgaIaThraValHisValAArgLeuLeuAaPaRgaAnaPaRPro	1020
Dp	3001	GCTCCTCTGGTGAACCCGGGTACAGTCCAGTCCGCTCTTGAACCGCAATGAACAACCA	3060
Qy	1021	ProValLeuGIaAaPheGIuIIeLeuPheAnaBaNTyrValThraAnaRgSerSerSer	1040
Dp	3061	CCAGTGGCGGCAACTTGAAGTCTTTCAACAACATATGCAACCATTCCTCAAGAGAC	3120
Qy	1041	PheProGIuGIaValaIIeGIaYgValProAlaHisAaPProAaPILeSerAaPSeu	1060
Dp	3121	TTCCCTGGGGGTGCATTTGGCCGAGTACCTGGCCCATGACCTGATATCTAGATAGTCTG	3180
Qy	1061	ThrTYrSerPheGIuAaRGIaAaGIaAaGIuLeuSerLeuValLeuLeuAnaIaSerThGIy	1080
Dp	3181	ACTTACAGCTTTGAGCGGGGAATATGACTAGCTGTGCTCTGCTCAATGCTTCCACGGGT	3240
Qy	1081	GIuLeuYrLeuSerAaRgaIaLeuAaPaRgaAnaPaRProLeuGIuValaIIeMeTyrVal	1100
Dp	3241	GAGCTGAAGCTAAACCCGCGCACTGACAAACAACGGGCTCTGGAGGCCATCATAGCGGTG	3300
Qy	1101	LeuValSerAaPgiYValHisSerValThraAGInCyAlaLeuAaGValThrIIeIle	1120
Dp	3301	CTGGTGTGACAGCGCGTACACAGGTGACCGGCCAGCGCGCGCTGTGACCATCATC	3360
Qy	1121	ThraAaPgiMeTyrThraHisSerIleThrLeuAaRgLeuGIuAaPMeTyrProGIuAaRg	1140
Dp	3361	ACCGATGAGATGCTCAACCAACAGATCACGCTGGCGCTGGAGGACATGTCAACCGAGCGC	3420
Qy	1141	PheLeuSerProLeuLeuGIuLeuPheIIeGInIaValaIaIaThrLeuIaIaThrPro	1160
Dp	3421	TTCTGTGTACCACTGCTAAGGCTCTTCTATCAGCGGGTGGCCGCGCACTGGCCACGCCA	3480
Qy	1161	ProAaPaRgaValValaIaPheAnaValGIaAaRgaPaRThraPaRAlaProGIuGIuHisIle	1180
Dp	3481	CCGGAACAAGTGGTGGTCTTCAACGTACAGCGGGAACAACGCCCCGGGGGGCGACATC	3540
Qy	1181	LeuAnaValSerLeuSerValGIuGIuProProGIuYProGIuGIuGIyProProPheLeu	1200
Dp	3541	CTCAACCTGAAGCCTGTGTGGTGGCGACGCCGCAAGGGCCGGGGGGCGCGCTTCTCTG	3600
Qy	1201	ProSerGIuAaPLeuGIuGIuAaRgLeuTYrLeuAnaRgSerLeuMeuThraIIaIleSer	1220
Dp	3601	CCCTCTGAAGACTGCAAGAGCGCTATACCTCAACCGCAAGCTGCTGAAGGCCATCTCG	3660
Qy	1221	AlaGIuAaRgValLeuProPheAaPaRgaAnaIIeCySeuAaRgGIuProCySGIuAnaTYr	1240
Dp	3661	GCAACAGCGCGTGGCTGCTTCAACGAACAATCTGCTGGCGGAGGCCCTGCAAGAACTAC	3720
Qy	1241	MeTyrGcyValSerValLeuAaRgPheAaPSeRerAlaProPheIIaIaSerSerSer	1260
Dp	3721	ATGGCGTGGGTGTGGGTGGCTGTGACTCTCTCGCGCTTCATCTCTCTCTCTCC	3780
Qy	1261	ValLeuPheAaPProIIeHisProValGIuGIuLeuAaRgCyAaRgCyPProProGIuPhe	1280
Dp	3781	GTGCTCTTTCGGGCCATTCACCCGCTGGAAGGGCTGGCTGGCCGGCCCGCGCTTC	3840
Qy	1281	ThrGIuAaPTRYGySGIuThrGIuValAaPLeuCySTyrSerAaRgProCySGIuProHis	1300
Dp	3841	ACGGGTACATATCGCGGAACCGAGGTGACCTGTGCTACTTCGGGCGCTGTGGCCCCAC	3900

QY	1301	GIYARGCYBAHGSERATGUGLIGLYGTYrThrCYBLEUCYBAHGAEPGLYrThrGLY	1320
Db	3901	GGGGGCTGCCACAGCCGGGAGGGCGGCTACACCTGCCCTCTCTCGTGAAGGCTACACGGGT	3960
QY	1331	GIWHASCSYGIUVALISERIALARGSERGLYARGCYSTRPARGLYVALCYBLVBAAGLY	1340
Db	3961	GAGCACTGTAGGTAGTGAATGCTCTCGTCAGGCCGTTGCACCCGGGTGTCTGGCAAGATGGG	4020
QY	1341	GIYrThrCYsVALISMLeuLeuVALIGLYIphelysCYsAAPCYsProSerGLYsAphe	1360
Db	4021	GGCAACTGTGTCAACCTCTGGTGGGGCGGTTCAAGTGCAGATTGGCCATTGGAGACTTC	4080
QY	1361	GIUlysProCYrCYsGLINVALThrThrArgSerPheProAlaHisSerPheIleThrThe	1380
Db	4061	GAGAAGCCCTACTGTGCAGGTGACACAGCGAGCTTCCCGGCTCACTCTTCATCACTCTT	4140
QY	1361	ARGGLYLeuARGLINARGPheHisApheThrLeuAlaLeuSerPheAlaThrLYSGLYARG	1400
Db	4141	CGGGGCGTGGCGCAAGCGTTTCCACTTCAACCTGGGCGCTCTGGTTGGCACAAGAGAGCG	4200
QY	1401	ASPGLYLeuLeuLeuTYrAsnGLYARGPheAsnGLYGLYHisPheVALAlaLeuGLU	1420
Db	4201	GACGGGTGTGCTGTGTATACATAGGGGGTTCATAGGAAGCATATGATTTGTGGCCCTCGAG	4260
QY	1421	VALIIEGGLINGLINVALIGLILEuThrPheSerAlaGLYLeuSerThrThrThrVALSer	1440
Db	4261	GTCATCCAGAGACAGGTCTCACTCACTTCTGTGAGGGAGATCACACACACGGGTGCC	4320
QY	1441	PROPheVALProGLYGLYVALISERASPGLYGINrPHIsthrVALIGLILEuLYsTYr	1460
Db	4321	CCATTCCGTGCCCGGAGAGATCAGTATGAGCGATGAGCATACGGTTCACCTGAATACTAC	4380
QY	1461	ASnLYsPROLeuLeuGLYGLINrThrGLYLeuProGINGLYPROSerGLINGLYsVALAla	1480
Db	4381	AATAAGCCACTGTGTGGTGTACAGAGGGCTCCACAGGGCCCATCAGACAGAAAGGTGCT	4440
QY	1481	VALVALThrVALASPGLYCYsAspThrGLYVALAlaLeuARGPheGLYSerVALLeuGLY	1500
Db	4441	GTCGTGACCGTGGAGTGGCTGTGACACAGAGATGGCTTGGCGTTGCGATCTGTCTGGGC	4500
QY	1501	ASnTYrSerCYsVALAlaGLINGLYrThrGLINGLYSerLYsPheSerLeuAspLeuThr	1520
Db	4501	AACrTACTCTGTGCTGCCACAGGGCCACCGAGGGTGGACAGAAATCTCTGTGATCTGACG	4560
QY	1521	GLYPROLeuLeuLeuGLYGLYVALProAspLeuProGLYsUserPheProVALArgMetARG	1540
Db	4561	GGGCCCCCTGTACTAGGGCGGGGTGCTTAACCTGCCGAGAGCTTCCAGTCCGAATGGCG	4620
QY	1541	GLINPheVALGLYsMetARGAsnLeuGLINVALAspSerATrHisILeAspMetALAAsp	1560
Db	4621	CAGTTCCGTGGGCTGTGATCGGGAACCTGCAGGGTGGACACCGGCACTATGACATGGCTGAC	4680
QY	1561	PHellLeuAlaAsnAsnGLYrThrVALProGLYCYsProAlaLYsLYsAsnVALCYsAspSer	1580
Db	4661	TTCArTTCGCAACAAAGGACCGGTGCTGTGGCTCTGTCCAGAAAGAGTGTGTGACAGC	4740
QY	1581	ASnThrCYsHISAsnGLYGLYrThrCYsVALAsnGLINrPAspAlaPheSerCYsGLYUCYs	1600
Db	4741	AACACTTCCACAAGTGGGGGCACTTGGGTGACACAGTGGAGACGCTTCACTCGACAGAGTCC	4800
QY	1601	PROLeuGLYpHeGLYGLYLYsSerCYsAlaGLINGLYMetVALAsnProGLNHISpHeLeu	1620
Db	4801	CCCCCGGGCTTTGGGGGCAAGAGCTGCCGCCACGGAAGGCGCAATCCACAGACTTCTCG	4860
QY	1621	GLYSerSerLeuVALAlaTrPHIAsGLYLeuSerLeuProILeserGLINrProTrPYrLeu	1640
Db	4861	GGCAGCAACCTGTGTGGCTGTGGCATGGCTTCTGTGCGCCACTTCCCAACCTGTGTACTTC	4920
QY	1641	SerLeuMetPheATrThrARGGLNAlaAspGLYVALLeuLeuGLINAlaIleThrARGGLY	1660
Db	4921	AGCTTCATGTTCCGACCGCGCAGGCGCACAGGTGTCTGTGCAGGCGCATCACAGGGGGG	4980

QY	1661	ArgSerThrIleThrLeuGlnLeuArgGluGlyHisIleMetLeuSerValGluGlyThr	1660
Db	4981	CGCAGCACCAATCAACCTTACAGCTACGAGAGGGCCACGTAAGTCTGAGCGTGGAGGCA	5040
QY	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnArgGlyAspTrpHis	1700
Db	5041	GGGCTTACAGGCTCTCTCTCTCCGTCTGGAGCCAGGCGGCGCATGACGTGGACTGGGAC	5100
QY	1701	HisAlaGlnLeuAlaLeuGlyValaSerGlyProGlyHisAlaAlaLeuSerPheAsp	1720
Db	5101	CATGACACGCTGGCACTGGGAGCCAGCGGGGGGCGTGGCCATGCAATTCGTCTCTTCGAT	5160
QY	1721	TyrGlyGlnGlnACGAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer	1740
Db	5161	TATGGGCAAGCAGAGCAGAGGAGGAACTTGGGCCCGCGCTGCATGCTTCACACTTGAAC	5220
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly	1760
Db	5221	AACTTAAACAGTGGGCGGAAATCACTGGGCGACCGCGGCTGTGGCCCGTTCGGGGCG	5280
QY	1761	CysLeuGlnGlyValArgValSerPheAspThrPheGluGlyValaAsnSerLeuAspProSer	1780
Db	5281	TGTTTGCAGGGGTGTGGCGGTAGCGAATACGCACAGAGGGGGTTAAACGCTGAACTCCAGC	5340
QY	1781	HisGlyGlnSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
Db	5341	CATGGGAGAGACATCAACGTGGACCAAGGGCTGTAGCTCGCTGACCTCTTGACTCAAC	5400
QY	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp	1820
Db	5401	CCGTGTCTGTCTAACGTATTTGACAGAACCACTGGGACAGCTATTCCTGACGCTGGAT	5460
QY	1821	ProGlyTyrTyrArgIleAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisGln	1840
Db	5461	CCAAGTTACTATGGTGAACACTGACTGACTGATGTGTGTAACCTGAAACCGGTGAGACACAG	5520
QY	1841	SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGluCysProProAsn	1860
Db	5521	TCTGTGTGTACCCCAAGCCCAAGCCGCCCCCATGACTTACCTCCAGTGTCCCCCAAT	5580
QY	1861	TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrTrpTrpGly	1880
Db	5581	TACCTTGGGCGCATCTGTGAGACAGATTTGACCAAGCCTTGTCCCCGTGGCTGGTGGGGA	5640
QY	1881	HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn	1900
Db	5641	CATCCACATGTGGCCCATGGCACTGTGATGTGACAAAGGCTTTGACCCGACATGCAAC	5700
QY	1901	LysThrSerGlyGluCysHisCysLysGluAsnHisGlyArgProProGlySerProThr	1920
Db	5701	AAGCAACGCGGAGTGCACATGCAAGAGAAACATACCGGCCCCAGGACGAGCCCCAC	5760
QY	1921	CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluLysP	1940
Db	5761	TGCCTCTTGTGTACTGTACCCCAAGGCTCTTGTCCAGAGCTGTGACCTCGAAGAT	5820
QY	1941	GlyGlnCysProCysLysProGlyValaIleGlyArgGlnCysAspArgCysAspAsnPro	1960
Db	5821	GGCCACAGTCCATCAACGACGAGTGTATCGGCGGTGACAGTGTGACCCGCTGTGAACACCT	5880
QY	1961	PheAlaGlyValaThrThrAsnGlyCysGlyValaAsnTyrAspSerCysProArgAlaIle	1980
Db	5881	TTTGCTGAGGTCACCAATGCTGTGAAATGTATATGACACTGCCCCACAGACGATT	5940
QY	1981	GluAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro	2000
Db	5941	GAGGCTGGGACTGTGTGGCCCCCGTACCCGCTTCGGGCTGACCTGTGCTGCCCTCTGCC	6000
QY	2001	LysGlySerPheGlyTyrThrAlaValaArgHisCysAspGlnHisAlaArgGlyTyrTrpLeuProPro	2020
Db	6001	AAAGGCTCTTTTGGACACTGTGGCGCACCTGTATATAGCAACAGGGGGTGGCTCCCCCA	6060
QY	2021	AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu	2040



```

Db      6061 AACCTTTCAAGCTGCAAGTCATCACCCTTCAGAACTAAGGCGCTTCGCTGAGCGGCTA 6120
Qy      2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060
Db      6121 CAGCGGAATGAGTCAGGCGCTAGACTCAGGCGCTCCAGAGAGTACCCCTGCTCCGCC 6180
Qy      2061 AamAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIysValAlaTyrGlnLeu 2080
Db      6181 AACGCCACCGACAGCACAGCTGCTACTTCGCGAGGAGCGTCAMAGGTGGCTTACCGAGCTG 6240
Qy      2081 AlaThrArgLeuLeuAlaHisGlnSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db      6241 GCCACGCGGCTGCTGCGCCAGAGAGACCCAGCGGCGCTTGGCGCTGTCGCACAGCAG 6300
Qy      2101 AspValHisPheThrGlnuAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db      6301 GACGTGCACTTCACCTGAGAAATGCTGCGGCGGTGGAGGCGCTCTCGAGCACAGCAAC 6360
Qy      2121 LysArgHisThrGlnLeuLeuGlnGlnThrGlnGlyGlyThrAlaThrLeuGlnHis 2140
Db      6361 AAGCGGCACTGGAGAGTGAATCCAGAGACAGAGGTGGACCGCTGCGCTCCAGCAC 6420
Qy      2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160
Db      6421 TATGAGGCTTACGCAAGTCGCTGCGCCAGAACATGCGGCACTCACTTAAGCCCTTC 6480
Qy      2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspIleGlyAsnPheAla 2180
Db      6481 ACCATGTCAGCGCCCAACATTGTCATCTCGTAGTGGCTTGGAACAAGGAACCTTGCT 6540
Qy      2181 GlyAlaIysLeuProArgTyrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200
Db      6541 GGGGCGCAAGCTCCCGCTACAGAGCGCTGCGTGGGAGACAGCCCGGAGACCTTGAACA 6600
Qy      2201 ThrValIleLeuProGlnSerValPheArgGlnThrProProValValArgProAlaGly 2220
Db      6601 ACGTCATCTTCGCTGAGTCTGCTGTCAGAGAGAGCGCCCGGTGTCAGGCGCGGAGCG 6660
Qy      2221 ProGlyGlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgArgHisPProGlnLeu 2240
Db      6661 CCGGAGAGGCGCCAGAGAGCCAGAGAGCGTGGACCGGCGACAGAGCGGACCGGAGCTG 6720
Qy      2241 SerGlnGlyGlnAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260
Db      6721 AGCCAGGAGGAGCGTGGCGAGCGTCATCATCCGACCCCTGCGCGCTACCTGCTCCT 6780
Qy      2261 HisAsnTyrAspProAspIysArgSerLeuArgValProIysArgProIleIleAsnThr 2280
Db      6781 CATTACTTATGACCTTGACAGAGCGAGCTTGAAGATCCCAAGCGCCGATCATCAACACA 6840
Qy      2281 ProValIleSerIleSerValHisAspAspGlnGlnLeuLeuProArgHisAlaLeuAspIys 2300
Db      6841 CCGGAGGAGAGCATCAGGCTCATGATGAGAGGCTTCTGCGCGCGCGCTTGACAA 6900
Qy      2301 ProValThrValGlnPheArgLeuLeuGlnThrGlnGlnArgThrLysProIleCysVal 2320
Db      6901 CCGGTCACGGTGCAAGTTCGCTGCTGAGACAGAGAGCGGACCAAGCCCATCTGTGTC 6960
Qy      2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGlyCysGln 2340
Db      6961 TTCTGGAACATTCAATCTTGATGAGTGCACAGGTGGCTGGCTGCGCACAGAGCGTGA 7020
Qy      2341 ValValPheArgAsnGlnSerHisValSerCysGlnCysAsnHisAsnMetThrSerPheAla 2360
Db      7021 GTCGCTTCCCGCAATGAGAGCACAGCTCAGCTGCAAGTGCACCATGACGAGCTTCGCT 7080
Qy      2361 ValLeuMetAspValSerArgArgGlnuAsnGlyGlnIleLeuProLeuLysThrLeuThr 2380
Db      7081 GTGCTCATGAGAGGTTTCTCGCGGAGATGGGAGATCTCTGCCATGAAGCACTGACA 7140
Qy      2381 TyrValAlaLeuGlnValThrThrLeuAlaAlaLeuLeuThrPhePheLeuThrLeu 2400

```

---

```

Db      7141 TACGTGGCTTAAGGTGTCACCTTGCGCTGCTTCGCTCACCTTCTTCTCACTCTC 7200
Qy      2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420
Db      7201 TTGGCTATCTGCGCTCCACCAACAGCGGATCCAGATCCATGACAGCTGCTCCGGGCG 7260
Qy      2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db      7261 CTGGCTCAGCTGCTTCTCTCTCGGGAATCAACAGAGCTGACCTCCCTTTCCTGCACA 7320
Qy      2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGln 2460
Db      7321 GTCATTTGCATCTCTGTGACTCTCTGTACCTGTGACCTTTTCTGAGGCTTGTGAGAG 7380
Qy      2461 AlaLeuHisIleTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480
Db      7381 GCTTGGACACTGTACCGGCACTCAAGAGTGGCGGATGTACAACCGGCCCATGCGC 7440
Qy      2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db      7441 TTCTACTACATGCTGGGCTGGGCGTGGCTGCTTCATCAACAGGCGTACGCGTGGGCTG 7500
Qy      2501 AspProGlnGlyTyrGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuLe 2520
Db      7501 GACCCGAGGCGTACGGAGACCTGACTTCTGTGCTGCTCTCATCTATGACAAGCTCATC 7560
Qy      2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db      7561 TGGAGTTTGTGCTGGCGCGGCTTTCGCTTCGCTGATGAGTGTCTTCTGTACATCTG 7620
Qy      2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnuLysGlyProValSer 2560
Db      7621 GCGGCGCGGCGCTCTGTGCTGCGCCAGCGGAGGCTTGGAGAAAGTCTGTCTG 7680
Qy      2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrThrPheLeuAlaLeu 2580
Db      7681 GCGCTGACGCGCTCTTCTGCGCTCTGCTGCTGAGACGCGCACGTGCTGCTGCACTG 7740
Qy      2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600
Db      7741 CTCTCTGTCACAGAGAACACCTCTCTTCACTACCTCTTGTGCACTGCAATTGCACTC 7800
Qy      2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerIysGlnValArgIysAlaLeu 2620
Db      7801 CAGGCGCCCTTCATCTCTCTCTCATGTGCTGTAGCAAGAGGTCCGGAAGAAGCACTC 7860
Qy      2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640
Db      7861 AAGCTTGCTGACAGCGGACCGGACCGCTGACCCCTGCTGTGACCAACAGTCCACCTG 7920
Qy      2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660
Db      7921 ACCTGCTCTTCAACTGCGCCAGCGCCCTACCGCAATGGCGGCTGTACACGCCCTTACGA 7980
Qy      2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680
Db      7981 GACTCGGCGGCTCTCTGACAGACCAAGTGTGCGGCAAGAGTCAAGCCAGCTACATC 8040
Qy      2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlnGlnIleProProGlyLeuGly 2700
Db      8041 CCCTTCTGTGAGAGAGAGATCCGACCTGAACCTTGCGCAAGGCGCCCTGCGCTTGCGG 8100
Qy      2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720
Db      8101 GATCAGAGCAAGCTGTTCTGTGAAGGTCAACACAGAGCATGATCTCGACACCGACTCC 8160
Qy      2721 AspSerAspLeuSerLeuGlnuAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740
Db      8161 GACAGTCACTGTCTTGAAGAGACACAGAGTGTCTCTTATGCTTACCACTGACATCA 8220
Qy      2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760
Db      8221 GACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280

```

QY 2761 TTPASPSerLeuLeuGlyProGlyValAlaGluArgLeuProLeuHisSerThrProLeuAsp 2780  
 Db 8281 TGGGATAGCTGCTGGGGCCCTGGAGCAGAGAGACTGCCCTCTGCACATGCTCCCAAGAT 8340  
 QY 2781 GAGGAGGCTGCTGGGGCCCTGGAGCAGAGAGACTGCCCTCTGCACATGCTCCCAAGAT 8400  
 Db 8341 GAGGAGGCTGCTGGGGCCCTGGAGCAGAGAGACTGCCCTCTGCACATGCTCCCAAGAT 8400  
 QY 2801 SerSerGlyAenGlyAlaProGluGluArgLeuArgGluAenGlyAenAlaLeuSerArg 2820  
 Db 8401 AGTAGTGGCAACGGGGCCCTGGAGCAGAGAGACTGCCCTCTGCACATGCTCCCAAGAT 8460  
 QY 2821 GluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisGlyGlyLeuLeu 2840  
 Db 8461 GAGGGGCTCCCAAGGCTCCCTCCAGGCTCTTCCAGCTCCCAAGGATCTCTTAAG 8520  
 QY 2841 LybLeuLeuValProHisLeuSerGlyLeuSerSerLeuLeuArgLeuProLeuGlu 2860  
 Db 8521 AAGAGTGTCTGCTCCACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8580  
 QY 2861 CysThrGlySerSerArgGlySerSerAlaSerGlyLeuArgGlyGlyProProPro 2880  
 Db 8581 TGCAACAGGCTCTTCCGGGGCTCTTCCGCTAGTAGAGGAGAGAGAGAGAGAGAGAGAG 8640  
 QY 2881 ArgProProProArgGlnSerLeuGlnGluLeuAenGlyValMetProLeuAlaMet 2900  
 Db 8641 CGCCCAAGGCTCCAG 8700  
 QY 2901 SerLeuLeuValGlyThrValAlaGlnLeuAenSerSerGlyLeuLeuPhePheAsn 2920  
 Db 8701 AGCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8760  
 QY 2921 PheLeuHis 2923  
 Db 8761 TTCCTGCAT 8769

RESULT 2  
 AX549238  
 LOCUS AX549238 10531 bp DNA linear PAT 26-NOV-2002  
 DEFINITION Sequence 523 from Patent WO02061087.  
 ACCESSION AX549238  
 VERSION AX549238.1 GI:25813930  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 Burner,G.C., Roush,C.L. and Brown,J.P.  
 Antigenic peptides, such as for G protein-coupled receptors  
 (GPCRs), antibodies thereto, and systems for identifying such  
 antigenic peptides  
 Patent: WO 02061087-A 523 08-AUG-2002;  
 Lifespan Biosciences, Inc. (US)

FEATURES  
 SOURCE  
 1. 10531  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 2038 a 3353 c 2998 g 2142 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 10531  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-916-849a-3 (1-2923) x AX549238 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 Db 63 ATCGAGAGCCCGGCGCACCGGGCGTCCCTCCCAACCGCGCGCGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyAenGlnValGlyProCysArgSerLeu 40  
 Db 123 TTCTCTCTGCTGCTGCGCGCGCACATTTGGAGAGACCAAGTGGGGCTCTCTCTCTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAlyAlyProMetGlyThrLeuLeuCysProSer 60  
 Db 183 GGGTCAGAGGAGAGAGAGGCTCTTCCGGGGCTGGGCGCCCAATGGGCTGGCTCTTCATCC 242  
 QY 61 SerAlaSerAenLeuThrLeuThrThrSerArgCysArgAenAlaGlyThrGluLeuThr 80  
 Db 243 TCAGGCTCCAACTCTGCGCTTACACAGCCGCTGAGGGATTCGGGCACTGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisLeuSerGlyLeuAenGlyValThrCysProGluSerGlyValHis 100  
 Db 303 GGCACCTGGATCCCAACCAACAGAGGCTGAGGGTTGGTGCAGAAATCCAGAGCCCAT 362  
 QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
 Db 363 ATTCCTCTACCAACAGCTCTTGAAGGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTG 422  
 QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGluGluHisProCysLeuVal 140  
 Db 423 GGCACCTTCCCAACAGGAGCAAGCTCACACTGCGAGAGAGAGAGAGAGAGAGAGAGAG 482  
 QY 141 ProArgLeuArgCysGlnSerCysLeuValAlaGlnAlaProGlyLeuLeuArgVal 160  
 Db 483 CAGGCTCAG 542  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgCysArgAenValAenThrAlaProGln 180  
 Db 543 AGGTCAACAG 602  
 QY 181 PheGlnProProSerGlyThrAlaThrValProGluAenGlnProAlaGlyThrProVal 200  
 Db 603 TTCACAGCCCGCGCTACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662  
 QY 201 AlaSerLeuArgAlaIleAenProAenGlnGluValAlaGlyValArgLeuGluThrMet 220  
 Db 663 GCATCCCTAGAGGAG 722  
 QY 221 AspAlaLeuPheAenSerArgSerAenGlnPhePheSerLeuAenProValThrGlyAla 240  
 Db 723 GATGCCCTCTTGTATAGCGGCTCCCAACAGAGTCTTCTCCCTGAGACCAAGTCACTGGGCA 782  
 QY 241 ValThrThrAlaGluGluLeuAenPheGlnThrThrThrThrThrThrThrThrThrThr 260  
 Db 783 GTAACCAACAGCCAG 842  
 QY 261 AlaGlnAenHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrThrThrThrThr 280  
 Db 843 GCGCAG 902  
 QY 281 AspThrAenAenHisAenProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
 Db 903 GACACCAATGACCATGACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
 QY 301 AenLeuGluValGlyThrGluValLeuThrValArgAlaThrArgPheValAenProPro 320  
 Db 963 AACCTGAGAGGTTGCTATAGAGGCTGCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1022  
 QY 321 AsnAlaAenIleLeuThrArgLeuLeuGluGlySerGlyGlySerProSerGluValPhe 340  
 Db 1023 AATGCCAATTTCTTATCCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082  
 QY 341 GluIleAenProArgSerGlyValIleArgThrArgGlyProValAenArgGluGluVal 360  
 Db 1083 GAGATGACCCCTGCTCTGAGGAGATCCAGAACCCGCTGCTGATCGGAGAGAGAGAGAGAG 1142  
 QY 361 GluSerGlyGlnLeuThrValGluValAenArgGlnGlyValArgAenProGlyProArgSer 380

Db	1143	GAATCTCTACGAGTCAACGGTAGAGCAAGTGAACGAGGTCGGGACCCGGGTCCTCGAGT	12027
Qy	381	ThrThrAlaAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPheSer	400
Db	1203	ACCAACGCGCGCTGTTTCTCTTCTGTGAGAGATGCAATGATTAAGCCCCCAAGTTTATG	12627
Qy	401	GluIysArgTyrValValGlnValArgGluAspValThrProGluValAProValLeuArg	420
Db	1263	GAGAAAGCGCTATGTGTGTCAAGGTGAGGAGAGATGTGACTCCAGGGGCCCAAGTACTCCGA	13222
Qy	421	ValThrAlaSerAspArgAspIysGlySerAsnAlaValValHisTyrSerIleMetSer	440
Db	1323	GTCAACGCTCGGATCGAGACACAGGGGAGCAATGCCGGTGTGCATTAAGACATCATGAGT	13822
Qy	441	GlyAsnAlaAcGgGlyGlnPheTyrTleuAspAlaGlnThrGluValAlaLeuAspValValSer	460
Db	1383	GGCAATGCTCGGGAGACAGTATTATTTGAGATGCCAGACTGAGAGCTGTGAATGTGTGAC	14422
Qy	461	ProIleuAspTyrGluThrThrTyrLeuGlyGluTyrThrLeuArgValArgAlaGlnAspGlyGly	480
Db	1443	CCTCTTGACTATGAGACGACCAAGGATGACCCCTAGGGGTGCAGACAGATGTGTGC	15022
Qy	481	ArgProProIleuSerAsnValSerGlyLeuValThrValGlnValLeuAspTlleAsnAsp	500
Db	1503	CGTCCCCACACTCTAATATGTCTGTGGCTGTGTGACAGTACAGGTCTCGGATATCAACAC	15622
Qy	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProIleu	520
Db	1563	AATGCCCCCATCTTGTCAGACACCCCTTCCAGGCTATGTGTCGAGAGGCTCCCTTA	16222
Qy	521	GlyTyrTleuValIleuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	1623	GGCTACCTGGTCTTCCAGTCCAGGCTACAGCGTATGCACGCTGATGCTGTGACAAAGCCCGCTG	16822
Qy	541	GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560
Db	1683	GAATACGCCCTTGCTGGGTGGAGATGACTTCCCTTCACATCAATCAAAATGGACAGGC	17422
Qy	561	TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal	580
Db	1743	TGGATCTCTGTGGCTGCTGAATCGGACGGGAGGAAGTTGATTCTACGCTTGGGGTA	18022
Qy	581	GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
Db	1803	GAAGCTCGAGACCAATGGCATTCACAGCACTGCTCGGCAAGTGCAGGTGATCTGTC	18622
Qy	601	LeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620
Db	1863	CTGATATGTCACAGCAACAAATCCAACTTACCCAAACAGAGTACACAGTGGCTCAAT	19222
Qy	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640
Db	1923	GAGAGTGCAGCTGTGGGACCAAGGTGTGACGGGTGTCACTGCTGTGACCCGTGATCTCAT	19822
Qy	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660
Db	1983	AGTGTATCATCTACCTACAGATACCAAGTGGCAATCTGCAAAACCGCTTCTCATCAACAC	20422
Qy	661	GlnSerGlyGlyGlyLeuValSerIleuAlaLeuProLeuAspTyrGlyLeuGluAlaArgGln	680
Db	2043	CAAGTGGTGTGGGCTGTATCCCTTGCCCTGCACTGAGACTCAAACTTGAACGGGACG	21022
Qy	681	TyrValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700
Db	2103	TATGTGTGTGGCTGTACCGCTCGAGTGGCACTGGGACGACACAGGCAATATGTGTG	21622
Qy	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720
Db	2163	AATGTCAACCGACGCAACCAACCATGTGCTGTCTTTCAAGACTCCCACTAATACGTGAT	22222
Qy	721	ValIleuGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740

Dp	2223	GTATATGAGGACCGGCGGCGAGGACCAACGGTGGTGGCTGATCAAGCGCCAGGATGAGGAC	2282
Qy	741	ThrgIygiuAenAlaargIleThrTyPheMetGluaspSerIleProGlnPheargIle	760
Dp	2283	ACAGGTGAGGAATGCCCGCATCACTACTTCAATGAGGAGCAGCAATCCCCCACTCCGCAATC	2342
Qy	761	AspAlaAspThrGlyAlaValaIleThrThcGlnalagIuAenAapTyrgIuaspGlyValaSer	780
Dp	2343	GATGAGAGCAAGGGGGGCTGTCAACCAAGGCTAGCTGAGCTACGAAAGACCAAGTGTCT	2402
Qy	791	TyrThreualaalllethralaargaspasnGlyIleProGlnySeraspThrThrTy	800
Dp	2403	TACACCCCTGGCCATTACTGTCTCGGGACAAATGGCAATCCCCAGAAAGTCCGACACCACTAC	2462
Qy	801	LeuGlnIleLeuValaenAapValaAspaspasnAlaProGlnIleleuAargaspSerTy	820
Dp	2463	CTGAGAGATCCCTGGGAGAACGAGTAAGAACAATGCCCTCCAGTTCCTGGCGAGCTCTCAC	2522
Qy	821	GlnGlySerValTyrgIuaspValaProProPheThrSerValleuGlnIleSerAlaThr	840
Dp	2523	CAGGGCAGGTGTCTATGAGGAGTGTCCACCCCTTCACTAGCGTCCGAGATCTCAGGCACCT	2582
Qy	861	AspArgaspSerGlyLeuaspnGlyAryValaPheTyThrPheGlnGlyGlyaspaspGly	860
Dp	2583	GATGTGTATCTTGAGACTTAATGGCAGGAGCTTCTTACACCTTCCAAAGAGGGGAGATGGA	2642
Qy	861	AspGlyaspPheIleValaGlySerThSerGlyIleValaArgThrLeuAargArgLeuasp	880
Dp	2643	GACGTGACTTTATTTGTGGAGTCCACGTCAAGCATCGTGGCAAGCTTACGAGGCTGGAT	2702
Qy	881	ArgGluasnValaIagIlnTyrgValleuAargAlaTyralaValaAapTyrgIyMetProPro	900
Dp	2703	CGAGGAACCGTGGCCGCAGTATGTCTTCCGGGCATATACAGTGGCAAGGGGATGCCCCCA	2762
Qy	901	AlaArgThrProMetGluValaIleThrValaThValleuaspValaAspaspasnProVala	920
Dp	2763	GCCCGCACCTTATGAGAGTGAACGTCACTGTGTGATGTGAATGACAAATCCCCCTGTC	2822
Qy	921	PheGlnIuAenApgIuAenAapValaPheValaGlnIuAenSerProIleGlyLeuAlaVala	940
Dp	2823	TTTGAAGCAGATGATGTTGATGTGTTGTGGAGAGAAAGACGCCCATTTGGGCTAGCCGTG	2882
Qy	941	AlaArgTyAlaThrAlaThrAapProaspGlnGlyThrasnAlaGlnIleMetTyrgIlnIle	960
Dp	2883	GCCCGGGGTCAAGCGCACTGACCCCGAATGAAGGCAACAAATGCCAGATTATGTACAAATT	2942
Qy	961	ValaGlnGlyAenIleProGluValaPheGlnleuAapIlePheSerGlyGluLeuThcAla	980
Dp	2943	GTGAGAGGGCAACATCCCTGAGGTCTTCACTGAGCAATCTTCTCCGGGAGCTCAAGACC	3002
Qy	981	LeuValaAapleuAapTyrgIuAaspArgProGlnTyrgValleuValaIleGlnAlaThrSer	1000
Dp	3003	CTGTGATGACTTGAAGCTACGAGAGACCGGCTGAGTAGCTCTCGGTCAATCCAGGCCACTCA	3062
Qy	1001	AlaProleuValaSerAargAlaThrValaHisValaArgleuAenAapArgAspaspasnPro	1020
Dp	3063	GCTCTCTGTGTACCCGAGCTCAAGTCAAGTCCGCTCTCTGACCGCAATGACAAACCA	3122
Qy	1021	ProValleuGlyAenPheGlnIleleuPheAaspasnTyrgValaThrasnAargSerSerSer	1040
Dp	3123	CCAGTGTGGGCAACTTTGAGATCTTTTCAACAATATGTCAACCAATCGCTCAAGCAGC	3182
Qy	1041	PheProGlyGlyValaIleGlyAargValaProAlaHisAapProaspIleSeraspSerleu	1060
Dp	3183	TTCCCTGGGGGTGCATATGGCCGAGTACCTGGCCATGACCTGATATCTCAAGATAGCTG	3242
Qy	1061	ThrTySerPheGlnAargIyAasnGluLeuSerleuValleuAenAlaSerThcGly	1080
Dp	3243	ACTTAAAGCTTTTGAAGGGGGAAATGAATCAAGCTGTCTGTGCTCAATGTCTCAAGGGT	3302
Qy	1081	GluLeuLeuValaenSerAargAlaAenAaspasnAapProleuGlnAlaIleMetSerVala	1100
Dp	3303	GAGCTGAAGCTTAACCCGCGACCTGGACAAACACCGGCTCTTGAGGGCATTCAATGAGGTG	3362

QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
DB 3363 CTGGGTCTGAGACGGCGGTACACAGCGTACCGCCAGCGCGCTGCGTGAACATCATTAC 3422  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
DB 3423 ACCGATGAGATGCTCACCCACAGCATCACGCTGCGCTGGAGGACATGTCAACCGAGCGC 3482  
QY 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaThrLeuAlaThrPro 1160  
DB 3483 TTCTCTGACACACTGGCTAGGCGCTTTCATTCAGCGCGGTGGCCGCCACGCTGGCCAGCCA 3542  
QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
DB 3543 CCGGACCACTGGTGGTCTTCAACGTACAGCGGACACCGACCGCCCGGCGGCACATC 3602  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPheLeu 1200  
DB 3603 CTCAAGCTGAGCGCTGTGCGTGGGCGAGCGCGCAGGCGCCGAGGCGCGCCCTTCCTG 3662  
QY 1201 ProSerGluAspLeuGlnArgLeuTyrlLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
DB 3663 CCTCTGAGGACCTGACAGAGCGCGCTTATCTTCAACCGCAGCCTGTGACGCGCATCTCG 3722  
QY 1221 AlAGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyrl 1240  
DB 3723 GCACAGCGCGTGGCTTCCCTTGCAGACACATCTGCTGCGGAGCGCCTGCGAAGACTAC 3782  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
DB 3783 ATGCGCTGCGTGTGCGTGGCGCTTGCATCTTCCGCGCCTTCATCGCCTCTCTCTCC 3842  
QY 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
DB 3843 GTGCTCTTCCGCGCCATCCACCCGTCGAGGCGTGGCTGCGCGCCCGCGCGCTTC 3902  
QY 1281 ThrGlyAspTyrlCysGluThrGluValAspLeuCysTyrlSerAlaProCysGlyProHis 1300  
DB 3903 ACCGGTGACTACTGCGAGACCGAGGTGAGCTCTGCTACTGCGCGCCTGTGCGCCAC 3962  
QY 1301 GlyArgCysArgSerArgGluGlyGlyTyrlThrCysLeuCysArgAspGlyTyrlThrGly 1320  
DB 3963 GGGCGCTGCGCAGCGCGAGGGGCGCTACCTGCTGTGTGATGCTACAGGCT 4022  
QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly 1340  
DB 4023 GAGCACTGTAGGAGTGTGCTGCTGCGCTGAGCCCGGTGTCTGCAAGATGGG 4082  
QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
DB 4083 GGCACTGTGTCAACTGCTGTGGGGGTTCGATTCAGATGCCATCTGAGACTTC 4142  
QY 1361 GlnLeuProTyrlCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
DB 4143 GAGAGCCCTACTGCGCAGGTGACACGCGAGCTTCCCGCCACTCTTCATCACCTTT 4202  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrGlyGlyArg 1400  
DB 4203 CGGCGCTGCGCGCAGCTTTCACCTTCACCTGCGCTCTGTTGCGCAAGAGAGCGC 4262  
QY 1401 AspGlyLeuLeuLeuTyrlArgGlyArgPheAsnGlyGlyHisAspPheValAlaLeuGlu 1420  
DB 4263 GACGGGTGTGTTGTTCATATGGCGTTTCATATGAGACATGACTTTGGGCGCTCGAG 4322  
QY 1421 ValIleGlnGlnValAlaGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
DB 4323 GTGATTCAGAGAGCGTCAAGCTCACTTCTCTGCGAGGAGTCAACACAGGATGCC 4382  
QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheThrValGlnLeuLeuTyrl 1460  
DB 4383 CCAATTCTGCGCGGAGGAGTCAAGTATGCGCACTGATCGGTGACGTGAGTAATACTAC 4442

QY 1461 AsnLeuProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLeuValAla 1480  
DB 4443 AATTAAGCCACTGTGGGTCAAGACAGGCTCCACAGAGGCCCATCAGAGCAAGAGTGGCT 4502  
QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
DB 4503 GTGGTGAACCGTGAAGGCTGTGTACACAGAGTGGCTTGGCTTGGATCTCTCCGGGC 4562  
QY 1501 AsnTyrlSerCysAlaAlaGlnGlyThrGlnGlySerTyrlSerLeuAspLeuThr 1520  
DB 4563 AACTACTCTGTGTGCTGCCAGGCGCACCCAGGGTGGCAGCAAGAAATCTCTGTGACTGACG 4622  
QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
DB 4623 GGGCGCCCTGCTACTAGGCGGGGTGCTTCACTGGCCGAGAGCTTCCAGTCCGAATGCGG 4682  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisAlaAspMetAlaAsp 1560  
DB 4683 CAGTTGTGGGCTGCAATGGGAACTGCAAGTGGACAGCGGACATACATGAGCTGAC 4742  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaGlyValAsnValCysAspSer 1580  
DB 4743 TTCAATTGCCAACAATGGACCGTGTGCTGGCTGCCCTGCCAAGAAACGTGTGTACAGC 4802  
QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGlyCys 1600  
DB 4803 AACACTTGGCAACAATGGGGCACTTGTGCGTGAACAGTGGAGACGCTTCAGCTGCGCAGTGC 4862  
QY 1601 ProLeuGlyPheGlyGlyCysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
DB 4863 CCCCTGGGCTTGGGGGAGAGCTGCGCCACAGAAATGGCCAAATCCACAGACTTCTCG 4922  
QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProThrTyrlLeu 1640  
DB 4923 GGACAGAGCTGTGTGCTGCGCATGAGCTCTGCTGCGCATCTCCCAACCTGATACCTC 4982  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
DB 4983 AGCCTCATGTTTCCGACCGCGCAGGCGCACGGTGTCTGTGCGACGCAATCACAGGGGG 5042  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
DB 5043 CGCAGACCACTACACCTTACGCTTACAGAGGCGCACGTGATCTGAGCGTGAAGGCGACA 5102  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
DB 5103 GGGCTTACAGCCTCTCTCTCTGCTGCTGAGGCGCAGGCGCGGCAATGACGTTGACTGCGAC 5162  
QY 1701 HisAlaGlnLeuAlaLeuGlyValAspGlyGlyProGlyHisAlaAlaLeuSerPheAsp 1720  
DB 5163 CATGACACACTGGCACTGGGAGCACGCGGGGGCGCTGGCCATGCAATCTGTCTTCGAT 5222  
QY 1721 TyrlGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisSer 1740  
DB 5223 TATGGCAGACAGAGCAGAGGGCAACTGGGCCCGCCCGCTGCAATGTGTGACCTGAGC 5282  
QY 1741 AsnIleThrValGlyGlyTyrlProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
DB 5283 AACATTAAGTGGGCGGAATACCTGGGCGAGCGGGGTGTGCGCGCTTGTGGGGC 5342  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
DB 5343 TGTTTGACAGGTGTGGGTGAGCAATAGCCAGAGGGGTTCATACAGCTGTGATCCAGC 5402  
QY 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
DB 5403 CATGGGAGAGCATCACTGAGAGCAAGCTTACGCTTACGACCTTGTGACTCAAC 5462  
QY 1801 ProCysProAlaAsnSerTyrlCysSerAsnAspThrAspSerTyrlSerCysSerCysAsp 1820  
DB 5463 CCGTGTCTGCTTAACAGCTAATGCGACAGACTGCGACAGCTAATCTCGCGACTGTGAT 5522  
QY 1821 ProGlyTyrlTyrlGlyAspAsnGlyThrAsnValCysAspLeuAsnProCysGluHisGln 1840

```
Db 5523 CCGAGTTACTATAGTGAACACTGACTAATGTGTGACCTGAACCCGTGTGAGCACAG 5582
Qy 1841 SerValCysThrArgLysProSerAlaProHisglYThrCysglYcysProProAsn 1860
Db 5583 TCTGTGTGATCCCGCAAGCCAGTGGCCCCCAAGGCTATACCTGGCGATGTCCCAAAAT 5642
Qy 1861 TyrLeuGlYProTyrCysglYThrArgLLeaspGlnProCysProArgglYTrpTrgLY 1880
Db 5643 TACCTGGGCGCATCTGTGAGACCAAGATTGACCAAGCTTGTCCCGGTGGTGGGA 5702
Qy 1881 HisProThrCysglYProCysAsnCyAspValSerLysglYPhesAppProAspCyAsn 1900
Db 5703 CATCCACAGTGGCCCATCACTGATGTCAGCAAGGCTTTTGACCCAGACTGCAAC 5762
Qy 1901 LysThrSerGlYglYcysHisCysLysglYAsnHisThrArgProProglYsAspProThr 1920
Db 5763 AAGACAGACGGCGAGTCCACTGCAAGAGAACCACTACGGGCCCCAGGACAGCCACC 5822
Qy 1921 CysLeuLeuCyAspCysTyrProThrGlYSerLeuSerArgValCyAspProGlYAsp 1940
Db 5823 TGCCTCTGTGTGACTGTACCCCAAGGCTCTTGTCCAGAGTCTGTGACCTGAGAGAT 5882
Qy 1941 GlYglYCysProCysLysProglYValLLeGlYArgGlncYAspArgCyAspAsnPro 1960
Db 5883 GGCACAGTGCATGCAAGCAGGAGTGTCTAGGCGTCAAGTGTACCGCTGTGACCAACCT 5942
Qy 1961 PheAlaGlYValThrThrAsnGlYcysGlYValAsnTyrAspSerCysProArgAlaLe 1980
Db 5943 TTTCGAGAGGTCAACACCAATGCTGTGAAGTAAATTATACAGCTGACCAAGAGAAAT 6002
Qy 1981 GlYAlaGlYLeTrpTrpProArgThrArgPheGlYLeuProAlaAlaLProCysPro 2000
Db 6003 GAGGCTGGAGTCTGTGGCCCGCTACCCGCTTGGGGCTGCTGCTGCTGCTGCTGCTG 6062
Qy 2001 LysGlYSerPheGlYThrAlaValArgHisCysAspGlYHisArgglYTrpLeuProPro 2020
Db 6063 AAGGCTCTCTTGGAGCTGTGTGGCGCACTGTGAATGAGACAGGGGCTGGCCCCCA 6122
Qy 2021 AsnLeuPheAsnCyThrSerLLeThrPheSerGlYLeuLysglYPhesAlaGlYArgLeu 2040
Db 6123 AACCTTTCACTGCACTGCACTGCTCACTTCTCAAGACTGAAGGGCTTCGCTGAGCGGCTA 6182
Qy 2041 GlYArgAsnGlYSerGlYLeuAspSerGlYArgSerGlncYLeuAlaLeuLeuArg 2060
Db 6183 CAGCGGAATGAGTCAAGGCTTGAAGTCAAGGCGCTCCAGCAAGCTACCCCTGCTGCGCC 6242
Qy 2061 AsnAlaThrGlnHisThrAlaGlYTrpPheGlYSerAspValLysValAlaTyrGlnLeu 2080
Db 6243 AAGCCACGAGACACAGCTGCTACTTCGAGCAGAGCTCAAGGTGAGCTTACAGCTG 6302
Qy 2081 AlaThrArgLeuLeuAlaHisglYSerThrGlnArgGlYPhesglYLeuSerAlaThrGln 2100
Db 6303 GCCACCGGCTGTGGCCCAAGAGACCAAGCGGGGCTTGGGCTGTCTGCAACAG 6362
Qy 2101 AspValHisPheThrGlnLeuLeuLeuLeuArgValGlYSerAlaLeuLeuAspThrAlaAsn 2120
Db 6363 GACGTCACTTCACTGAGAACTGTGCTGGGAGTGGGAGCGGCTCTTGGACACAGCAAC 6422
Qy 2121 LysArgHisThrGlnLeuLLeGlnGlnTrpTrgLYglYThrAlaTrpLeuLeuGlnHis 2140
Db 6423 AAGCGGCACTGGAGGTGATTCAGACAGACAGAGGTGGCACCGCTGTGCTCCAGCAC 6482
Qy 2141 TyrGlYAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160
Db 6483 TATGAGCGCTAGCGCAGTGGCTGGCCAGAGACATGCGGACACCTTACAGCCCTTTC 6542
Qy 2161 ThrLLeValThrProAsnLLeValLLeSerValValArgLeuAspLysglYAsnPheAla 2180
Db 6543 ACCATGTGTACGGCCAACTGTCTCATCTCGTGTGAGCAAAAGGAACTTGTGCT 6602
Qy 2181 GlYAlaLysLeuProArgTyrGlYAlaLeuArgGlYGlYLeuProProAspLeuGlnThr 2200
Db 6603 GGGGCCAAGCTGCCCCGCTACGAGGCTTGGTGGGAGACAGCCCCGAGACTTGAACA 6662
Qy 2201 ThrValLLeLeuProGlYSerValPheArgGlYThrProProValValArgProAlaGlY 2220
Db 6663 AAGATATTCTGTGATGTGTGTCTTCAAGAGAGACCCCCGTGTGAGGGCCCGAGGC 6722
Qy 2221 ProGlYGlYAlaGlnGlYProGlYGlYLeuAlaArgArgGlYArgArgHisPProGlYLeu 2240
Db 6723 CCGAGAGAGGCCAGGAGCCAGAGAGAGCTGGCACGGCGACAGGACGCGACCCGAGCTG 6782
Qy 2241 SerGlnGlYGlYAlaValAlaSerValLLeLLeTyrArgThrLeuAlaGlYLeuLeuPro 2260
Db 6783 AGCCAGGAGTGAAGCTGTGGCCAGCGTCATCATCTACCGCACCTGTGGCCGAGCTGCT 6842
Qy 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProLLeLLeAsnThr 2280
Db 6843 CATTAATAAGACCTTGAACAGCGCAGCTTGAAGATGCCAAAGCCCATCATCAACA 6902
Qy 2281 ProValSerLLeSerValHisAspAspGlYGlYLeuLeuProArgAlaLeuAspLys 2300
Db 6903 CCGTGTGAGATCAGCTGCATATATATGAGAGCTTCTGCCCGGGCCCTGAGCAAA 6962
Qy 2301 ProValThrValGlnPheArgLeuLeuGlnTrgLYglYArgThrLysProLLeCysVal 2320
Db 6963 CCGGTCACGGTCAAGTCCGCTGCGTGGAGACAGAGAGCGGACCAAGCCATCTGTGTC 7022
Qy 2321 PheTrpAsnHisSerLLeLeuValSerGlYThrGlnGlYTrpSerAlaArgLysCysglY 2340
Db 7023 TTCTGAACTTCAATCTGTGATGAGTGCAGAGTGGTGTGGCGGAGAGGCTGTGA 7082
Qy 2341 ValValPheArgAsnGlYSerHisValSerCysGlnCysAsnHisAsnMetSerPheAla 2360
Db 7083 GTGCTTCCGCAAGAGAGGACAGCTCAGCTGCCAGTGCACATGACAGACTTCGCT 7142
Qy 2361 ValLeuMetAspValSerArgArgGlYLeuGlnLLeuProLeuLysThrLeuThr 2380
Db 7143 GTGTCTATGAGAGTTCCTCGGGGAGAAATGGGAGATCTGCCACTGAAGACACTGACA 7202
Qy 2381 TyrValAlaLeuGlYValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400
Db 7203 TACGTGGCTTGAAGTGTCACTTGGCTGCTGCTTGTGCTCACTTCTTCTTCACTCTC 7262
Qy 2401 LeuArgLLeLeuArgSerAsnGlnHisglYLLeArgArgAsnLeuThrAlaAlaLeuGlY 2420
Db 7263 TTGCGATTCCTGCGTCCAAACCAACAGGCAATCCAGCACTGACAGCTCCCTGGGCT 7322
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlYLLeAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db 7323 CTGGCTCAGCTGTCTTCTCTCGGAAATCAACAGGCTGACCTCCCTTTTGGCTGACA 7382
Qy 2441 ValLLeAlaLLeLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGln 2460
Db 7383 GTCATTTGCCATCTCTGTGCACTTCTGTACCTTGCACCTTTTCTGGGCTGTGCTGAG 7442
Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlYProMetArg 2480
Db 7443 GCGTTGACACTGTACCGGCACTCACTGAGGAGCGGAGTGTCAACACCGGCCCATGGGC 7502
Qy 2481 PheTyrTyrMetLeuGlYTrpGlYValProAlaPheLLeLLeThrGlnLeuAlaValLLeu 2500
Db 7503 TTCTACTACATGCTGGGCTGGGGCTGTGCTGCTTCAACACAGGCTAGCCGTGGGCTG 7562
Qy 2501 AspProGlYGlYTyrGlYAsnProAspPheCysTrpLLeSerLLeTyrAspThrLeuLLe 2520
Db 7563 GACCCCGAGGCTACGGGAACCTTGACTTCTGTCTGTCTTCACTTATGACAGCTCATTC 7622
Qy 2521 TrpSerPheAlaGlYProValAlaPheAlaValSerMetSerValPheLeuTyrLLeLeu 2540
Db 7623 TGAATTTTGTGGGCGCGGTGGCTTTGCGCTGAGTGAAGTGTCTTCTGTACATCTG 7682
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlYPheGlnLysLysglYProValSer 2560
Db 7683 GCGGCCGCGGCTCTGTGTGCGCCAGCGGAGGGCTTTGAGAAAGAGTCTGTCTCG 7742
```

Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
 Db 7743 GGCTGGAGGCTCTCTGCGGCTCTCTGCTGAGGCGCAGTGGCTGCTGCACTG 7802  
 Qy 2581 LeuSerValaAnSerAapThrLeuLeuPheH1sTryLeuPheAlaThrCysAanCysIle 2600  
 Db 7803 CTCTCTGTCAACACGACGACCTCTCTCTCACTCACTCTCTCTCTCTCTCTCTCT 7862  
 Qy 2601 GlnGlyProPheH1sPheLeuSerTrpValValLeuSerLysGluValArgValAlaLeu 2620  
 Db 7863 CAGGAGCCCTTCACT 7922  
 Qy 2621 LysLeuAlaCysSerArgLysProSerProAapProAlaLeuThrTrpLysSerThrLeu 2640  
 Db 7923 AAGCTTGCTGCAAGCCGAGGAGCCAGGCTGAGCTGCTGAGCAGCAGTCACTG 7982  
 Qy 2641 ThrSerSerTrpAanCysProSerProTrpAlaAapGlyValGlyLeuTrpGlnProTrpGly 2660  
 Db 7983 ACCCTGCTTCAACACTGCGCCAGGCTCTCAAGATGGCGCTGTACCGAGCCCTACGGA 8042  
 Qy 2661 AapSerAlaGlySerLeuH1sSerThrSerArgSerGlyLysSerGlnProSerTrpIle 2680  
 Db 8043 GACTCGGCGGCTCTCTGCAAGCAGAGCTGCTGCGCAAGAGCTCAGCCAGCTCACTC 8102  
 Qy 2681 ProPheLeuLeuAArgLysLysAlaLeuAanProGlyGlnGlyProProGlyLeuGly 2700  
 Db 8103 CCTCTCTGCTGAGGAGGAGGAGCTCGCACTGAGCCTGCGCCAGGAGGCGCCCTGCGTGGG 8162  
 Qy 2701 AapProGlySerLeuPheLeuGlnGlyGlnAapGlnGlnH1sAapProAapThrAapSer 2720  
 Db 8163 GATCAGGAGGAGCTGCTCTGGAAGGTCAAGACGAGCTGATCTCTGACACGAGCTCC 8222  
 Qy 2721 AapSerAapLeuSerLeuGlnAapAapGlnSerGlySerTrpAlaSerThrH1sSerSer 2740  
 Db 8223 GACCTGACCTGTCTCTTGAAGACGACGAGTGGCTCTCTTACCTCACTCACTCACTCA 8282  
 Qy 2741 AapSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 Db 8283 GACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8342  
 Qy 2761 TrpAapSerLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2780  
 Db 8343 TGGGATAGCTGCTGCGGCTCTGAGCAGAGGAGTGCCTCTGACAGTCACTCCCAAGAT 8402  
 Qy 2781 GlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 2800  
 Db 8403 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8462  
 Qy 2801 SerSerGlyAanGlyAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2820  
 Db 8463 AGTAGTGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8522  
 Qy 2821 GlnGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2840  
 Db 8523 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8582  
 Qy 2841 LysLysCysLeuProThrIleSerGlnLysSerSerLeuLeuAArgLeuProLeuGlnGln 2860  
 Db 8583 AAGAGGTCTCTGCGCAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8642  
 Qy 2861 CysThrGlySerSerArgLysSerSerAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2880  
 Db 8643 TGCAAGAGGCT 8702  
 Qy 2881 ArgProProAaArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2900  
 Db 8703 CGCCACCGCCCGGAGAGGCTCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8762  
 Qy 2901 SerIleLysAlaGlyLysValAapGlnAapSerSerGlySerGlnPheLeuPhePheAan 2920  
 Db 8763 AGCATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8822

Qy 2921 PheLeuH1s 2923  
 Db 8823 TTCCTGAT 8831  
 RESULT 3  
 AF234887  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 10531)  
 Vincent, J.B., Skaug, J. and Scherer, S.W.  
 The human homologue of flamingo, *Sox12*, encodes a brain-expressed  
 large cadherin-like protein with epidermal growth factor-like  
 domains, and maps to chromosome 1p13.3-p21.1  
 DNA Res. 7 (3), 233-235 (2000)  
 MEDLINE  
 PUBMED  
 10907856  
 2 (bases 1 to 10531)  
 Vincent, J.B., Skaug, J. and Scherer, S.W.  
 Direct Submission  
 Submitted (15-FEB-2000) Department of Genetics, The Hospital for  
 Sick Children, 555 University Avenue, Toronto, ON M5G 1X6, Canada  
 JOURNAL  
 TITLE  
 AUTHORS  
 REFERENCES  
 MEDLINE  
 PUBMED  
 10907856  
 2 (bases 1 to 10531)  
 Vincent, J.B., Skaug, J. and Scherer, S.W.  
 Direct Submission  
 Submitted (15-FEB-2000) Department of Genetics, The Hospital for  
 Sick Children, 555 University Avenue, Toronto, ON M5G 1X6, Canada  
 FEATURES  
 source  
 1. .10531  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p21.1"  
 /clone\_lib="Stratagene fetal brain library, Cat. number  
 H13003a"  
 63. .8834  
 /note="FMI; brain-specific, cadherin-like protein; similar  
 to the Homo sapiens product encoded by clone KIA0279 in  
 GenBank Accession Number D87469 and Rattus norvegicus  
 M8692 encoded by GenBank Accession Number AB011528"  
 /codon\_start=1  
 /product="FLAMINGO 1"  
 /protein\_id="AAG00080.1"  
 /db\_xref="GI:9828190"  
 /translation="MSSPATGVPLPPPPILLLLLLPPLAGDVGPCRSISRG  
 KSSGACAPWMLCPSSASNLMTYRSCDAGTEULGHLVPHHDGRVPCRSBAHP  
 LPPAPGCPWCKLGLIGHLSPQGLTLPBEPCLAKPRLCQSCKLQAPGLRAGE  
 RSPBSLIGRRKRNVTAPQFPQPSYQAVPBNQAGTPVASIRALDPDEGEARLEY  
 TMDALFDSRNPFPGLDPTGAVTTEBLRERKTHFRVTAQDHGMPRRSALATLT  
 ILVTNDNDPVEBOEYKSLRENLVEYGLVTAQDGPAPNNILYRLLEGSGG  
 SPSEVEIDPREGVTRGPDREERESYQVTAQDGRDGRDGRSTAAVPLSVSD  
 NDNAAPSEKRYVYQVREYVPGAPVLTARTASDRDGSNAVHYTSMGNARQPLD  
 AQTGALDVSPLEETTKLVRKQDGERPLSNVSGLVTVQVLDINDNAPITVST  
 PFGATLVSPLEGLTVLHVALDADGDNARLEYLAGVGHPEPTIINGTMI SVAA  
 ELDEREVDYFSGEARHDGTPALTASASVSVVLDVNDNPPFTQPEYTRLMEDA  
 VGSVTVVAVDRAEDAVITVYQTSNTNRFSITQSGGGLVSLPLDYRLKERYV  
 LAYTASDGTROPAQIVVAVTANTHRPVFQSHYVYNNEDRPAGTIVLLISAEEDT  
 TGNMAITTPMEDSIPRIPIDATGAVTTOABLDYEDOVSYTLAITARPNGIPKSDT  
 TTEILVNDVNDAPQPLDSTQGSYEDVPEFTSTQISATDRDGLNARVYTTQG  
 GDGDGDFIVESTSGIVRLRLDRENVQVYLAAYADKMPATPMEVTVYLDV  
 NDNPVFBQDEPDVFBENSPIGLAVARTADPDGTNAQVQIVBQNIPEVQLD  
 IFAGELTALVDLYEDRPERVAVTQATSAPLVSRAVHVALDRNDNPPVLAGNEILF  
 NNVTNRSSFPGLAIGRPADHPDIDSILTSYFSEGNELSLVLAASNGELLSAL  
 DNRPILBAIMSVYSDGVSVAQCLARTYITIDEMLTSTILRLDMSBRLSLD  
 GLTQAVATLATPDPHYVYVNYQRTDAPGHTILIVASISVQPPRPGGSPRLPSBD  
 LGRLTLNLSLTLAISAQVLPDPDNI CLAREPENTMRCSVAFPSAPFLSSVY  
 FRILHPVGLRCRCPGFTADYCEITVDLCYSPCPHGCRRBGGYTCCLDGTYG  
 EHECVARSGRCPPGVCKNGGTCVNLVSGFKDCSDGPEKCYCVVTRSPFASHT  
 TTPGRHORHFTALSFATKRDGILLNNGRFRKDKDLYALVIOVOVLTFSAGST  
 TTVSPVPGVSDGWHVTLQRYNNKPIILGQGLPGGSRBOKYAVTVVGGCDTGVALR  
 FGSVLNYSCAAQGTGSGSKSLDLTGLPILLAGVPPDLPSPFVRMQPFCGRANLYVD



SRHIMADPIANNNGTVRCGPARKVNDSTNGCTCVQNDAPFSCCECPYAGRGKSCA  
QEMANPQFTLSSGLVAMHGLSLPISQWYLSLMEFRQADGVLLQATIGRSTITLQ  
REGWMLSVETGLQASLSLRLPRANDMDWHQALQALASGSGPHALISPYGOORA  
BENLQPRHLGLHLSNITVGGIPGAGVAGRGGLQGVADNTEGVSLDPSSGES  
INVEQGLSLPDCSDNCPGANSYCSNDMDYSGSCDPPYGNCTNVCDINCEHOSV  
CTKRSAPHGTYCECPNYLGPYCETRIDOPCRGMWHPCTGCKMVCYKSPDPDN  
KTSQCHCKENYRPPGSPCTCLCCCTPFGSLSRVCDPDDCCPCPGVIGQCRCD  
NPPAVYTNCEVNNDSCPRALBAGIWPRTPLPAAPCRKSGFVAVRQDBRRG  
WLPPLNCTSLITFSELKGFARLQDNESGLDSGSQDALLRNATQHTAGYFGSDV  
KVAQDLATRLASLSEBORGFLSATQDVFTEMLRVGSLDITANKRHEWLIQOTEG  
GTAMLLQHYEASALAQNRHTYLSPTITVPTNIVSVRLDKNAPGAKLPRBAL  
RSEOPDLTETVILPESEVRETPVVRPGBOPEABOPEALROKRNHPLSGEIVAS  
VIVRYLGLALPHANDPDRSRARVPRPIINTPVVISIYHDBELLPRALDPVVOF  
RLLETRERTKPICVFMNHSILVSGTMSNARCEVFNRESVSCQCHMTSFAYLMD  
VSRRENGELPLKTLTYVALGVTLAALLTFPFLTLRLIRSNONGIRNLTALGLA  
QVPLIGINQADLPACTVIALILHFLYCTSPWALLLEHLRYALTEVRDNTGPMR  
FLYLMGWVPAFTGLAVLDEBEGNPFQWLSYDITLINSFAGVAVAMSELY  
ILAAASCAAOBGPEKGVGVSGLQSPFVLLLSATWLLALSNSDPLPHYVLPAT  
CNCTGSPRIELSYVYLSKEVRKALKACRKSPPPALTTKSTLISYRCPYADGR  
LYQPTGDSAGSLHSTSRGSKQPSYIPLIRESALNPGQGPGLADPESLLEQDQ  
QHPDPTSDSLSLDDQSGSYASTHSDBSEEBEERBAAPBOGDWLSLGGAB  
RLPHSTPDGPGPKAPWPDFTTAKESGNGAPERLRENDLALSREGLPLP  
GSSAQBPKGILLKKCLPTISEKSLRLPLEQCTGSSRSGSSASBGRGPPRP  
SLQEDNGMPTAMSLKAGTVEDSGSEFLFPNPLH"

BASE COUNT 2038 a 3353 c 2998 g 2142 t

## ALIGNMENT SCORES:

Pred. No.: 0 Length: 10531  
Score: 2923.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-916-849a-3 (1-2923) x AF234887 (1-10531)

QY 1 MetAArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
DB 63 ATGCGGAGCCCGGCCACCGCGCTCCCTCCCAACCCCGCCGCTGCTGCTGCTG 122  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAserGlnValGlyProCysAArgSerLeu 40  
DB 123 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
QY 41 GlySerArgGlyAArgGlySerSerGlyAlyCysAlaProMetGlyTyrLeuCysProSer 60  
DB 183 GGGTCCAGGGGAGAGAGGCTCTTCCGGGGGCTGCGCCCAATGGGCTGGCTCTGCTCATTC 242  
QY 61 SerAlaSerAserAserLeuTyrLeuTyrThrSerArgCysAArgAserAlaGlyThrGluLeuThr 80  
DB 243 TCAGGCTCCAGACCTGCTGCTCATACCAAGCCCTGCAAGGATGCGGGCACTGAGCTGACT 302  
QY 81 GlyHisLeuValProHisAserGlyLeuAArgValTyrCysProGluSerGlnAlaHis 100  
DB 303 GGCACCTGAGTACCCCAACAACAATGAGGCTTGGTGTTCAGAAATCCAGAGCCCAT 362  
QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysAArgLeuLeuGlyTyrIleGly 120  
DB 363 ATTCCTCCATACCAAGGCTCTGAGAGGCTGCCCTGAGAGTGTGCTCTCTGGGCAATTGGA 422  
QY 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGlnHisProCysLeuLeuVal 140  
DB 423 GGCACCTTTCCCAACAAGGCAAGCTCACATGCCCCGAGAGCAACCCGCTTAAAGGCT 482  
QY 141 ProArgLeuArgCysGlnSerCysLeuLeuAlaGlnAlaProGlyLeuAArgAlaGlyGlu 160  
DB 483 CCAAGGCTCAAGTCCAGTCTGCAAGCTGGCAAGGCCCCCGGGCTCAAGGCAAGGGA 542  
QY 161 ArgSerProGluGlnSerLeuGlyGlyAArgGlyAArgAserValAserThrAlaProGln 180  
DB 543 AGGTACACCAAGAGTCCCTGGGTGGGCTCGGAAAGGAAATGTAATACAGCCCCCAG 602

QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAserGlnProAlaGlyThrProVal 200  
DB 603 TTCAGCCCCCAAGTACCAAGGCCACAGTCCGAGAACCAAGCCAGCGAGCACTCCCTGTT 662  
QY 201 AlaSerLeuAArgAlaIleAserProAserGlyGlyValAArgValAArgLeuGlyTyrThrMet 220  
DB 663 GCATCCCTGAGAGGCCATGACCCCGGACAGAGGTGAGGAGGTGACCTGAGATCACCATG 722  
QY 221 AserAlaLeuPheAserSerAserAserGlnPhePheSerLeuAserProValThrGlyAla 240  
DB 723 GATGCCCTCTTGTGATAGCCGCTCCACCAAGTCTTCTCTCGAGACCAAGTCACTGAGTGA 782  
QY 241 ValThrThrAlaGluGlnLeuAserAArgGlyLeuThrIleAserThrHisValPheAArgValThr 260  
DB 783 GTAACCAAGCCAGGAGGATCGATGATGAGACCAAGACCAACCAAGTCTTCAAGGCTCAAG 842  
QY 261 AlaGlnAserHisGlyMetProAArgAArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 843 GCGCAGACCAAGGATCCCGGACGAGAGTCCCTGAGCTACCACTCAATCTTGGTACT 902  
QY 281 AserThrAserAserHisAserProValPheGluGlnGlnGlyTyrIleGlyIleSerLeuAArgGlu 300  
DB 903 GACACCAATGACATGACATCCCTGCTGAGCAGCAGAGATCACAGAGACCTCAGGAG 962  
QY 301 AserLeuGlnValGlyTyrGlnValLeuThrValAArgAlaThrAserGlyAserAlaProPro 320  
DB 963 AACCTGAGAGTGGCTGATGAGGTGCTCACTGACAGGCAAGATGATGATCCCTCC 1022  
QY 321 AserAlaAserIleLeuTyrAArgLeuLeuGlnIleSerGlyGlyIleSerProSerGlyValPhe 340  
DB 1023 AATGCAATATTCGTACCGCTGCTGAGAGGCTGAGGAGGCTGAGGAGGCTGAGAACTCTTT 1082  
QY 341 GluIleAserProAArgSerGlyValIleAArgThrAArgGlyProValAserAArgGluVal 360  
DB 1083 GAGATGACACCTCCCTGCTGAGGATGATCCGAAACCCGTGCTGATCCGGAAGAGGTG 1142  
QY 361 GluSerTyrGlnLeuThrValGluAlaSerAserAserGlnIleAArgAserProGlyProAArgSer 380  
DB 1143 GATCTCTACAGCTGACGATGAGGAGGACAGTACCAAGGCTCGGAGACCCGGGCTCTGAGAT 1202  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAserAserAserAserAserAserAserAser 400  
DB 1203 ACCACAGCGCTGCTTCTCTTCTGAGGAGATGACATGATGATGATGATGATGATGATGAT 1262  
QY 401 GluIleAserGlyValValGlnValAArgGluAserValThrProGlyValAProValLeuAArg 420  
DB 1263 GAGAGGCTCATGTGTGCTGACAGTGAAGGAGATGATGATCCAGGGGCCCAAGTACTCCGA 1322  
QY 421 ValThrAlaSerAserAserAserAserGlySerAserAlaValIleAArgSerIleMetSer 440  
DB 1323 GTCAAGGCTCGATGACAGACAGGAGGACATGCCGTGAGTCACTATGATCATGAGT 1382  
QY 441 GlyAserAlaAArgGlyGlnPheTyrLeuAserAlaGlnThrGlyAlaLeuAserValValSer 460  
DB 1383 GGCATATGCTCGGAGACAGTATTATCTGATGATCCCAAGCTGAGCTGAGATGTGTAGGC 1442  
QY 461 ProLeuAserTyrGlnThrThrIleGlyTyrThrLeuAArgValAArgAlaGlnAserGlyGly 480  
DB 1443 CCTCTGATCATGAGACGACCAAGAGATGACCTTACCGGCTGCGACAGCAAGATGTGTGC 1502  
QY 481 ArgProProLeuAserAserAserGlyLeuValThrValGlnValLeuAserIleAserAser 500  
DB 1503 GGTCCCCCACTCTATATGTCTCTGAGTGTGATGACAGTACAGTCTGATATATCAACAGC 1562  
QY 501 AserAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
DB 1563 AATGCCCCCATCTTCTGTCAGACACCTCTTCCAGGCTACTGTCTGAGAGGCTCCCTTA 1622  
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAserAlaAserAlaGlyAserAserAserAser 540  
DB 1623 GGCTACCTGGTTCATCATGCTCAGGCTATGAGAGCTGATGTGATGATGATGATGATGATGAT 1682  
QY 541 GluTyrAArgLeuAlaGlyValGlyHisAserPheProPheThrIleAserAserGlyThrGly 560





QY 1281 ThrGlyAspTyrCysglutThrValAspLeuCysTyrSerArgProCysglYProHis 1300  
 Db 3903 ACCGGTGAACCTACTGACGAGACCGAGGTGACCTCTGCTACTCGGGCCCTGGCCCCCAG 3962  
 QY 1301 GlyArgCysArgSerArgGluGlyYrYrThrCysLeuCysArgAspGlyYrYrThrGly 1320  
 Db 3963 GGGCGCTGGCCGACGGCGGACGGCGGCTACCTGCTCTGTGTATGATGCTACAGGGGT 4022  
 QY 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysLysAsnGly 1340  
 Db 4023 GAGCACTGTGAGGTGAGTGTCTCCCTAGGCCCTTGCACCCCGGGTGTCTGCAGAAATGGG 4082  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACCTGTGTCAACCTGCTGTGGCGGGTTTCAAGTGGCATTTGCCCATCTGGAGACTTC 4142  
 QY 1361 GlyLysProTyrCysGluValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 Db 4143 GAGAAACCCCTACTGACGAGGTGACACGCGACGCTCCCGCCCATCTTCATCACCTTT 4202  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
 Db 4203 CGGGGCTGCGCAGCGCTTCCACTTCACTGCGCTCTCTGTTCCCAAGAGAGCCG 4262  
 QY 1401 AspGlyLeuLeuLeuTyrArgGlyArgPheAsnGlyLysHisAspPheValAlaLeuGlu 1420  
 Db 4263 GACGGGTGTCTGTGTATCAATGAGGCGCTTCAATGAGAGATGATGCTTGTGGCCCTCGAG 4322  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLysThrThrThrValSer 1440  
 Db 4323 GTGATTCAGAGAGAGGTCCAGCTCACCTTCTGCGAGGGAGTCAACACACGCGGTGCC 4382  
 QY 1441 ProPheValProGlyValSerAspGlyGlnThrPheIleThrValGlnLeuLysTyr 1460  
 Db 4383 CCATTGCTGCCCGAGAGAGTCACTATGCGCACTGGCATACGGTGCAGCTGAATCTAC 4442  
 QY 1461 AsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnLysValAla 1480  
 Db 4443 AATAAGCCACTGTGGGTGACAGAGGGCTCCCAAGGGCCCATCAGAGCAAGAGGTGCT 4502  
 QY 1481 ValIleValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 Db 4503 GTGGGTACCGGTGATGCTGTGTACACAGAGAGTGGCTTGTGCTCGGATCTGTCTCGGGCC 4562  
 QY 1501 AsnTyrSerCysAlaIleAsnGlyThrGlnGlyGlySerLysGlySerLeuAspLeuThr 1520  
 Db 4563 AACTACTCTGTGCTGCCAGGGCACCCAGGGTGGCAGCAAGAGTCTTGATCTTGACG 4622  
 QY 1521 GlyProLeuLeuLeuGlyValProAspLeuProGlyLysPheProValArgMetArg 1540  
 Db 4623 GGGCCCTCTACTAGCGGGGGTCTGACTGCCGAGAGCTTCCCACTCCGAATGGCG 4682  
 QY 1541 GlnPheValGlyCysMetLysArgAsnLeuGlnValAspSerLysGlnIleAspMetLysAsp 1560  
 Db 4683 CAGTGTGTGGGTGTGACGTGGAACTGACGTGGACACCGCGCACTAACATGTGTGAC 4742  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysGlyAsnValCysAspSer 1580  
 Db 4743 TTCAATGGCCACATGGACCGTGTGCTGCTGCTGCCAAGAAACGTGTGTGACACG 4802  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnTTPAspAlaPheSerCysGluCys 1600  
 Db 4803 AACACTGTGCACAAATGGGGCACTTGTGTGAACAGATGGGACGCGTTGACGTGGAGTGC 4862  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
 Db 4863 CCCCTGGGCTTGTGGGGCAAGAGCTGTGCGCCAGAAATGCAATCCAGACACTTCTCG 4922  
 QY 1621 GlySerSerLeuValAlaThrPheGlyLysSerLeuProIleSerGlnProTyrPheLeu 1640  
 Db 4923 GGCAGAGAGCTGTGGCTGTGGATGGCTCTGTGCTGCCATCTCCCAACCTGTGTACTTC 4982

QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 4983 AGCTTCATGTTCCGACCGCGCCAGCGGTGCTCTGTCTGACAGGCATTCACAGAGGGG 5042  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlyGlnHisValMetLeuSerValGluGlyThr 1680  
 Db 5043 CGCAGACCATCAACCTCAAGCTACAGAGGGCCACGTGATCTGACGTGTGAGGGCACA 5102  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 Db 5103 GGGCTTCAGGCTCTCTCTCTCTGAGCCAGGCCCGGGCCAAATGACGTATCTGGCAC 5162  
 QY 1701 HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 5163 CATGCACAGCTGGCATCTGGAGCCAGCGGGGGCTGTGGCAATGCCATTCGTCTTCGAT 5222  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLysLeuSer 1740  
 Db 5223 TATGGGACAGAGAGAGAGAGGCAACTGGGCCCGCGCTGTGCATGTCTGCACCTGAGC 5282  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 Db 5283 AACATACATGTGGCGGAAATCTGGCCAGCGCGGGGTGTGGCCGTGTGACTTGGGGC 5342  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValaSerLeuAspProSer 1780  
 Db 5343 TGTTTGACAGGGGTGTGGGTGAGCATACGCCAGAGGGGGTTTACAGCTGTGATCCAGC 5402  
 QY 1781 HisGlyGlyLysSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 Db 5403 CATGGGAGAGATCAATCAAGTGGAGCAAGCTGTACCTGTGACCTTGTGATCACTAAC 5462  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyrSerCysSerCysAsp 1820  
 Db 5463 CGGTCTCTGTACAGCTATTTGACGACACACATCTGGGACAGTATTTCTGACCTGTGAT 5522  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840  
 Db 5523 CCAAGTACTATGTGGCAACTGTACTATGTGTGTGACGTGAACCGGTGTGAGCAGCAGC 5582  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 Db 5583 TCTGTGTATCCCGCAACGCCAGTGGCCCCCATGTGCTTACTGTGAGATGTGCCCAAT 5642  
 QY 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrTyrPheGly 1880  
 Db 5643 TACTTGGGCAATACTGTGAAACAGAGTTGACAGACTTGTCTCCCGTGGCTGTGGGGA 5702  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn 1900  
 Db 5703 CATCCCATGTGGCCCATGTGCACTGTGATGTGACAAAGGCTTTGACCCAGACTGCAAC 5762  
 QY 1901 LysThrSerGlyGlyLysCysHisCysLysGlyLysAsnHisGlyTyrArgProProGlyLysProThr 1920  
 Db 5763 AAGACAAAGCGGAGTGTCACTGTGACAGAGAAACCACTACCGGCCCGGAGGAGCCCAAC 5822  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyLysAsp 1940  
 Db 5823 TGCCCTTGTGTGACTGTACTCCCAAGAGCTCTCTTGTCCAAAGTGTGTGACCTTGAGAT 5882  
 QY 1941 GlyLysCysProCysLysProGlyValIleGlyValArgLysCysAspArgCysAspAsnPro 1960  
 Db 5883 GGCAGTGTCCATGACAGCCAGAGTGTGATCGGGGCTCAGTGTGACCGGTGTGACAAACCT 5942  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle 1980  
 Db 5943 TTGTGTAGGTACACCAATGAGCTGTGAATGAAATTAAGACGTGTGCCACAGAGCGATT 6002  
 QY 1981 GluAlaGlyIleTyrTyrProArgThrArgPheGlyLeuProAlaAlaIleProCysAspPro 2000  
 Db 6003 GAGGTGGAGATCTGTGGTCCCGGTATCCCGCTTGGGCTGTGCTGTGCTCTCTGTGCC 6062  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlyHisAspGlyTyrPheLeuProPro 2020

```

Db      6063 AAAGGCTCTTGGAGCTGCTGGCCGACCTGAGAGCAGAGGGGTGGCTCCCCCA 6122
Qy      2021 AsnLeuPheAsnCySerThrSerIleThrPheSerGluLeuIleuValPheAlaIleuArgLeu 2040
Db      6123 AACCTTCACTGACGACGCTCATCACTTCTCAGAACTGAAGGGCTTCGCTGAGCGGCTA 6182
Qy      2041 GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnIleuValLeuLeuLeuArg 2060
Db      6183 CACGGGAATGAGTCAGGCTAGACTCAGGCGCTCCACGACGATAGCCCTGCTCTGGCG 6242
Qy      2061 AsnAlaThrGlnIleSerThrAlaGlyTyrPheGlySerAspValIleValAlaIleArgIleu 2080
Db      6243 AACGCCACGAGACACAGCTGCTACTTGGCAGCAGCAGCTCAAGGTGGCTTCCAGCTG 6302
Qy      2081 AlaThrArgLeuLeuAlaIleGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db      6303 GCCACGGGCTGCTGGCCCAAGAGACCCACAGGGGGCTTTGGGCTGTCTGCCACACAG 6362
Qy      2101 AspValIleAspThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db      6363 GACGTGCACTTCACTGAGATCTGCTGGGTGGGCAAGCCCTCTCCAGCACAGCCCAAC 6422
Qy      2121 LysArgHisThrProGluLeuIleGlnIleThrGluGlyTyrAlaIlePheLeuGlnHis 2140
Db      6423 AACCGGCACTGGAGCTGATCCAGCAGACAGAGGTGGCACCGCTGGCTGCCAGCAC 6482
Qy      2141 TyrGluAlaIleTyrAlaSerAlaLeuAlaGlnAspMetArgHisThrTyrLeuSerProPhe 2160
Db      6483 TAGAGGCTTACGCCAGTGCCTGGCCCAAGACATGGGCAACCTACCTAGACCTTCC 6542
Qy      2161 ThrIleValIleThrProAsnIleValIleSerValIleArgLeuAspIleGlyAsnPheAla 2180
Db      6543 ACCATCGTCAAGCCCAACATTTGTCATCTCGTAGTGGCTTGGCAAAAGGGAATTTGCT 6602
Qy      2181 GlnAlaIleLeuProArgTyrGluAlaLeuArgGlyGluIleProProAspLeuGluThr 2200
Db      6603 GGGGCGAAGCTGCCCCGCTACAGAGGCTCTGTGGGAGAGAGCCCGGACCTTGAACA 6662
Qy      2201 ThrValIleLeuProGluSerValPheArgGluThrProProValIleArgProAlaGly 2220
Db      6663 ACGATCTTGTGCTGAGTGTGTCTTCAAGAGACGCCCCCGTGTACAGCCCGAGAGC 6722
Qy      2221 ProGlyGluAlaGlnIleProGluIleuValIleArgArgGlnIleArgArgHis 2240
Db      6723 CCGGAGAGGCGCCAGGACAGAGAGCTGACGAGGACAGGACCGGACCGGAGCTG 6782
Qy      2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuIleGlyLeuLeuPro 2260
Db      6783 AGCCAGGGGTAGGCTGTGGCCAGGCTCATCTACCGCACCCCTGGCCGGGCTAGCTGCT 6842
Qy      2261 HisAsnTyrAspProAspIleAspSerLeuArgValIleProIleAspProIleIleAsnThr 2280
Db      6843 CATACATATGACCTCTGACAGAGGAGCTTGAAGTCCCAACGCCCGACATCAACACA 6902
Qy      2281 ProValIleSerIleSerValIleAspAspGluIleuLeuProArgAlaLeuAspIle 2300
Db      6903 CCGCTGTGAGCATCAGCGTCAATGATGAGGAGCTTCTGCCCGGCGCTCGACAAA 6962
Qy      2301 ProValIleThrValGlnPheArgLeuLeuGluThrGlnIleuArgThrIleuAspProIleVal 2320
Db      6963 CCGCTCAAGGAGCTTCCGCTGCTGAGACAGAGAGCGAGACCAAGCCCATCTGTGTC 7022
Qy      2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGly 2340
Db      7023 TTTCGGAACCATTCATCTGTGTCACTGGGCAAGGTGGCTGTGGCCAGAGGCTGTGAA 7082
Qy      2341 ValValIlePheArgAsnGluSerHisValSerGlyGlnIleAsnHisMetThrSerPheAla 2360
Db      7083 GTCTCTCTTCGCAATGAGAGCAGCTCAGCTGCCAGAGCAACCAATGAGAGCTTGTCT 7142
Qy      2361 ValLeuMetAspValIleSerArgArgGluAsnGlyGluIleuProLeuValThrLeuThr 2380

```

```

Db      7143 GTCTCATGAGCGTTTCTCGCGGAGAAATGGGAGATCTGCCACTGAAACACTGACA 7202
Qy      2381 TyrValAlaLeuGlyValIleThrLeuAlaIleLeuLeuThrPhePhePheLeuThrLeu 2400
Db      7203 TAGGTGGCTTACAGGTGTACCTTGGCTGCTTGTCTGCTCACTTCTTCTCTCACTCTC 7262
Qy      2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIleuGly 2420
Db      7263 TTGGCTATCTTCGCTCCACCAACAGCGCATCCCAAGTAACTGACAGCTCCCTGGGCG 7322
Qy      2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCySerThr 2440
Db      7323 CTGGCTCAAGTGTCTTCTCTCGGGAATCAACAGGCTGACCTCTTGTCCCTGCA 7382
Qy      2441 ValIleAlaIleLeuLeuHisAspLeuTyrLeuCySerThrPheSerTyrAlaLeuGln 2460
Db      7383 GTCATTTGCATCTGTGACCTCTGTATCCTCTGACCTTTCTCGGGCTGTGGGAG 7442
Qy      2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValIleArgAspValAsnThrGlyProMetArg 2480
Db      7443 GCTTGGACCTGTACCGGCACTCACTGAGGTGGCCGATGTCAACACCGGCCCATGGCGC 7502
Qy      2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db      7503 TTCTACTACATCTGGGCTGGGCGCTGCTGCTTCAATCAAGGGCTACCGTGGGCTG 7562
Qy      2501 AspProGluGlyTyrGlyValAsnProAspPheCySerTyrLeuSerIleTyrAspThrLeuIle 2520
Db      7563 GACCCGAGGGGTAGGGGAACCTGACTTGTGTGCTGCTGCTCATATGACAGCGTCATC 7622
Qy      2521 TTPSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db      7623 TGGAGTTTGTGCGGCGCGGTGGCTTTGGCTGTGAGAGTGTCTTCTGTACATCTCG 7682
Qy      2541 AlaAlaArgAlaSerCyAlaAlaIleGlnArgGlnIlePheGluLysValGlyProValSer 2560
Db      7683 GGGGCGGGGCTCTGTGCTGCTCCAGGCGGAGGCTTTGAGAAAGAGTCTGTCTG 7742
Qy      2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaIleThrTyrLeuLeuAlaLeu 2580
Db      7743 GGCCTGAGGCTCTCTGTGCGCTGCTGCTGCTGAGGCGCACGAGGTGTGGCACTG 7802
Qy      2581 LeuSerValAsnSerAspThrIleuLeuPheHisTyrIleuPheAlaIleThrCyAsnCyIle 2600
Db      7803 CTCTGTCTCAACAGGACACCTCTCTTCACTACCTCTTGTCACTGCAATTCATC 7862
Qy      2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerIleGlyValIleArgValAlaLeu 2620
Db      7863 CAGGGCCCTTCACTTCTCTCTCTATGTGTGTGTAGCAGAGAGGTCCGAAAGCATCTC 7922
Qy      2621 LysLeuAlaCySerArgIleAspProAspProAlaLeuThrThrIleSerThrLeu 2640
Db      7923 AAGCTTGCTGTGACGCGGAGCCAGCCCTGACCTGCTGACCAACCAAGTCCACCTG 7982
Qy      2641 ThrSerSerTyrAsnCySerProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660
Db      7983 ACCTGCTCTTACAACTGCCCCAGCCCTTACCGAGATGGCGGCTGTACCAAGCTTACGGA 8042
Qy      2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680
Db      8043 GACTCGGCGGCTCTCTGACAGCAGCACTCCCTCGGGAAGATCAAGCCAGCTACATC 8102
Qy      2681 ProPheLeuLeuArgGluIleuSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700
Db      8103 CCTTCTTGTCTGAGGAGAGTCCCACTGAACCTGTGGCAAGGGGCCCCCTGCGGGG 8162
Qy      2701 AspProGlySerLeuPheLeuGluIleGlnIleAspGlnIleAspProAspThrAspSer 2720
Db      8163 GATTCAGGAGGCTGTCTTGGAGGTCAAGACAGACAGCAAGATCTTCAACAGGACTCC 8222
Qy      2721 AspSerAspLeuSerLeuGluIleuAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740
Db      8223 GACAGTGACTGTCTTAAAGACAGCAGAGTGGCTCTTAAGCCCTTACCACTCATCA 8282

```



Qy 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1081 GAATTCACAGCGACGCGTGAAGGCAAGTGAACAGGGTGGGACCCGGGTCTCCGAGT 1140  
 Qy 381 ThrThrAlaValPheLeuSerValGluAspAspAspAspAspAspAspAspAsp 400  
 Db 1141 ACCAAGCCCGCTGTTCTCTTCTGTGAGGATGACATATATATATATATATATATATAT 1200  
 Qy 401 GluLysArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db 1201 GAGAGGCGCTATGTGTGTCAGGTGAGGAGATGTGCTCCAGGGGCCCAAGTATCCGA 1260  
 Qy 421 ValThrAlaSerAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 440  
 Db 1261 GTCAAGCCCTCGATTCAGACAGAGGAGGACATCCCTGGTGCATATACATCATAGT 1320  
 Qy 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 Db 1321 GGCAATGCTCGGGGACAGTTTATCTGATGCCCCAGCTGAGCTGATGTGTGAGC 1380  
 Qy 461 ProLeuAspTyrGlnThrThrLysGlnTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
 Db 1381 CCTTTACATAGACAGACAGACAGAGATACACCTACGGGTGCGAGCAAGAGATGTGTC 1440  
 Qy 481 ArgProProLeuSerAspAspValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1441 CGTCCCACTCTTATGCTCTGCTGCTGGTGAACATACAGCTCCGATATACAGAC 1500  
 Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 Db 1501 AATGCCCCCACTCTGTCAGACACCCCTTCCAGGCTACTGCTCCGAGAGACGTCCTTCA 1560  
 Qy 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db 1561 GGCTACTGCTGTTCACATGTCAGAGCTATGACGCTATGCTGTGCAATATGCCGCTG 1620  
 Qy 541 GlnTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAspGlyThrGly 560  
 Db 1621 GAATACGCGCTGTGCTGGGTGGGACATGACTTCCCTTCCACATCAACATATGACAGGC 1680  
 Qy 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
 Db 1681 TGGATCTCTGTGTGCTGTAAGTGAACCGGAGGAGATGTATTTCTACAGCTTGGGGTA 1740  
 Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerValSerValThrVal 600  
 Db 1741 GAAGCTGAGACATGGCACTCCAGCACTCACTGCTCGCAGATGTCAGGATGACTGTC 1800  
 Qy 601 LeuAspValAlaAspAspAspProThrPheThrGlnProGlnTyrThrValArgLeuAsn 620  
 Db 1801 CTGATGTCAACGACCAACATCCAACTTTACCAACAGAGTACACAGTGGCGCTCAAT 1860  
 Qy 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAlaAspArgAspAlaHis 640  
 Db 1861 GAGGATCAGCTGTGGGACACGAGGTGTGACGCTGTGACGCTGTGACCGTGAAGCTCAT 1920  
 Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 1921 AGTGTACATCACTACAGATACAGAGTGGCAATATCGAAGACCGCTCTCCATACACAGC 1980  
 Qy 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLysGluArgGln 680  
 Db 1981 CAAGTGGTGTGGGCTGTATCCCTTCCCTGCACTGACATCAAACTTGAACGGCAG 2040  
 Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgIleAspThrAlaGlnIleValVal 700  
 Db 2041 TATGTGTGTGTGCTTACCGCTCCGATGGCACTGGGACAGACAGGACAGATTTGTGTG 2100  
 Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrValAsn 720  
 Db 2101 AATGTACCGACGCGCAACCAACCATGCTCTTCTTTCAGAGCTCCCACTATACAGTAAAT 2160

Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db 2161 GTTAAATAGGACCGCGCGGACAGCAACAGGTGTGTCTATACAGGCCCAAGATAGAGAC 2220  
 Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 Db 2221 ACAGGTGAAGATCCCGCATCACTTCAATGAGAGACAGATCCCGCATGTCGCAATC 2280  
 Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnAlaSer 780  
 Db 2281 GATGACAGACCGGGGCTGTACACACAGCTGAGCTGAGCTGACAGAACCAAGTGTCT 2340  
 Qy 781 TyrThrLeuAlaIleThrAlaArgAspAspGlyIleProGlnLysSerAspThrThrTyr 800  
 Db 2341 TACACCTGGCCATTAAGTCTGGGACATGSCATTCCCAAGATCGACACCACTTAC 2400  
 Qy 801 LeuGlnIleLeuValAlaAspAspValAlaAspAspAspAspAspAspAspAspAsp 820  
 Db 2401 CTGGAGATCTGTGTGAACACATGATGACAAATGCCCTCCAGCTTCTGCGAGACTCTTAC 2460  
 Qy 821 GlnGlySerValTyrGluAspValProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2461 CAGGCGAGTGTCTATAGGATGTGCCACCTTCACTAGCTGCTGACAGATCTCAGCACT 2520  
 Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2521 GATCGTGAATTCGAACTTAATGGCAGGGCTCTTACACTTCCCAAGAGCGCAGATGGA 2580  
 Qy 861 AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 Db 2581 GACGGTGACTTTATGTTATGTTAGTCAAGTCAAGGACATCGTGCAGACGTACGAGGCTGAT 2640  
 Qy 881 ArgGluAspValAlaGlnTyrThrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900  
 Db 2641 CAGAGAACCTGGCCCAAGATGTCTTGGGCAATATGACATGACAGAGGAGATGCCCA 2700  
 Qy 901 AlaArgThrProMetGlnValThrValThrValLeuAspValAlaAspAspProVal 920  
 Db 2701 GCCCGCACACTATGAGATGACATGCTGATGTGATGATGATGATGATGATGATGATGAT 2760  
 Qy 921 PheGlnGlnAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAlaVal 940  
 Db 2761 TTTGACAGAGATGATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 2820  
 Qy 941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 2821 GCCCGGTACAGCCCACTGACCCCAATGAGGACCAACGCCCAATATATATATATATATAT 2880  
 Qy 961 ValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980  
 Db 2881 GTGAGGCGCAACATCCCTGAGGCTTCCAGCTGACATCTTCCCGGAGGCTGACAGCC 2940  
 Qy 981 LeuValAspLeuAspTyrGluAspArgProGlnTyrThrValLeuValIleGlnAlaThrSer 1000  
 Db 2941 CTGTGAGCTTGAATGACATGAGAGACCGGCTGAGTACGTCCTGTCAATCAAGGACAGTGA 3000  
 Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspAspPro 1020  
 Db 3001 GCTCTCTGTGTAGCGCGGCTACAGTCCAGCTCCGCTCTTGAACCGCAATGACACCA 3060  
 Qy 1021 ProValLeuGlyAsnPheGlnIleLeuPheAsnAspTyrValThrAsnArgSerSer 1040  
 Db 3061 CCAAGTCTGGGCAACTTGTAGATCTCTTTCACCAATATATGTCACCAATGCTCAACACAC 3120  
 Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 3121 TTCCTGGGGGTGCCATTTGCGAGTACCTGGCCAGACCCGTGAATCTCAGATATGCTG 3180  
 Qy 1061 ThrTyrSerPheGlnArgGlyAsnGlnLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3181 ACTTACAGCTTTGAGGGGGAATGAACTACAGCTGTGTCTGTCAATGCTTCAAGGGGT 3240  
 Qy 1081 GluLeuLysLeuSerArgAlaLeuAspAspAspArgProLeuGlnAlaIleMetSerVal 1100

Db 3241 GAGCTGAAGCTTAAGCGCGGCACTGGACAACACCGGCTCTGGAGGCGCATCAAGAGGTG 3300  
Qy 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1120  
Db 3301 CTGGGTGCACAGCGCGTACACAGGCTGACCGCCAGTGGCCGCTGCGTGGACCAATC 3360  
Qy 1121 ThrAspGlnMetLeuThrHisSerIleThrLeuArgLeuGlnAspMetSerProGluArg 1140  
Db 3361 ACCGAATGAGATGCTACCCACAGCATCAACCTGGCGCTGGAGGACATGTCAACCCGAGCGC 3420  
Qy 1141 PheLeuSerProLeuLeuGlnPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
Db 3421 TTCTCTGACACACTGCTAGGCTCTTTCATCAAGCGGTGGCCGCAAGCTGGCCAGCGCA 3480  
Qy 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
Db 3481 CCGGACACAGTGTGGTCTTCAACGTACAGCGGGAACCGAACCCCGCGGGGCCCAATC 3540  
Qy 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu 1200  
Db 3541 CTCACAGTGAAGCTGTGTGGTGGCCAGCCCGCAGGGCCCGGGGGCGGCTCTTCCTG 3600  
Qy 1201 ProSerGlnAspLeuGlnGlnArgLeuThrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
Db 3601 CCTCTGAGGACCTGGACGAGCGCTATACCTCAACCGCAGCTGTGCAGCGCATCTCG 3660  
Qy 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGlnProCysGlnAsnThr 1240  
Db 3661 GACACAGCGGTGCTGGCTTTCAGACAAACATCTGCTGGGGAGCCCTCGAGAACTAC 3720  
Qy 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
Db 3721 ATGCGGTGGGTGTGGGTGGCTTGCATCTCCGCGCGCTTTCATCGGCTCTCTCC 3780  
Qy 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280  
Db 3781 GTGCTTTTCGAGCCATCCACCCCGTGGAGGCTGGCTGGCCGCGCCCGGCGCTTC 3840  
Qy 1281 ThrGlyAspThrCysGlnThrGlnValAspLeuCysArgSerArgProCysGlyProHis 1300  
Db 3841 ACGGAGTACTAGTCCAGACCGAGGTGGACTCTGCTGACTCGGGCCCTGGCCCCCA 3900  
Qy 1301 GlyArgCysArgSerArgGlnGlyGlyThrCysLeuCysArgAspGlyThrGly 1320  
Db 3901 GGGCGGTGCGCAGCGCGGAGGGCGGCTACACTGCTGTGTATGCTACAGGGGT 3960  
Qy 1321 GlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGly 1340  
Db 3961 GAGCATGTGAGGTGAGTGTCTCCCTAGCGCGGTGCAACCCGGGTGTGTGCAAGATGG 4020  
Qy 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
Db 4021 GGGACTGTGTCAACTGCTGTGGGGCGGTTCATATGGCATTTGCCCATCTGGAGCTTC 4080  
Qy 1361 GlnIleProThrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
Db 4081 GAGAAACCCCTACTCGCAGGTGACACAGCGAGCTTCCCGCCCATCTTCATCACTTT 4140  
Qy 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleArg 1400  
Db 4141 CGGGGCTGGCGCAGGCTTTCACATCACTCGGCTCTCGTTTCCCAAGAGAGCGC 4200  
Qy 1401 AspGlyLeuLeuLeuThrArgGlyArgPheAsnGlnIleHisAspPheValAlaLeuGln 1420  
Db 4201 GACGGGTGTCTGTGTACATGGCGCTTCAATGAGAGATGACTTTGTGGCCCTCGAG 4260  
Qy 1421 ValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440  
Db 4261 GTGATTCAGAGAGAGGTTCAGCTCACTTCTGCGAGGGAGTCAACCAACAGGTGTCC 4320  
Qy 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuIleThr 1460

Db 4321 CCATTCGTGCCCGGAGAGTCAAGTAGCCAGTGGCATACGGTGCAGCTGAATATCTAC 4380  
Qy 1461 AsnIleProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnIleValAla 1480  
Db 4381 AATTAAGCCATGTGGGTGATGACAGGGCTTCCACAGGGCCCATCAAGACAGAAAGTGGCT 4440  
Qy 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4441 GTGTGACCGGTGATGTGCTGTGACACAGAGTGGCTTCCGCTTGGATCTGTCTGGGGC 4500  
Qy 1501 AsnIleThrSerCysAlaAlaGlnGlyThrGlnGlyGlySerIleGlySerLeuAspLeuThr 1520  
Db 4501 AACTACTCTGTGTGTGCCAGGGACCCAGGGGTGGACAGCAAGATCTGTGATTCAGAG 4560  
Qy 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGlyLeuSerPheProValArgMetArg 1540  
Db 4561 GGGCCCTCTCTACTAGCGGGGTGCTTACCTGGCCGAGAGCTTCCCACTCCGAATGGCG 4620  
Qy 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
Db 4621 CAGTTCGTGGCTGTGATGGGAACCTGGCAGGTGACAGCGGCACATAGACATGGCTGAC 4680  
Qy 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIleGlyAsnValCysAspSer 1580  
Db 4681 TTCATTTGCCAACAATGGACCGGTGCTGGCTGCTGCGCCACAAAGAACGTGTGTACAGC 4740  
Qy 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
Db 4741 AACATTTGCCACAAATGGGGGCACTTGTGTGAAACAGGTGGAGCGCTTCACTGCGAGTCC 4800  
Qy 1601 ProLeuGlyPheGlyGlyIleYsSerCysAlaGlnGlnMetAlaAsnProGlnHisAspLeu 1620  
Db 4801 CCCCTGGGCTTTGGGGGCAAGAGCTGCCGCCAGAAATGGCAATCCACAGCATTCCTG 4860  
Qy 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProThrPheLeu 1640  
Db 4861 GGCACAGCGTGTGTGCTGTGATGCTCTGCTGCGCCATCTCCCAACCTGTGTACCTC 4920  
Qy 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCTTCATGTTCCGACCGCGCAGGCCAGCGGTGTCCTGTGCAAGCATCCAGGGGG 4980  
Qy 1661 ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGlyThr 1680  
Db 4981 CGCACACACATCAACCTACAGCTACAGAGGGCCACGTGATGCTGAGGTGGAGGGCA 5040  
Qy 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspThrHis 1700  
Db 5041 GGGCTTCAGGCTCTCTCTCTCGTGTGAGCCAGGCCGAGCCAAATGACGTGACTGGCAC 5100  
Qy 1701 HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
Db 5101 CATGCACAGCTGGCACTGGAGGCCAGCGGGGGCCTGGCCATGCTGTGCTTCAT 5160  
Qy 1721 TyrGlyGlnGlnAspGlnGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
Db 5161 TATGGGCAAGAGAGCAGAGGGCAACTGGGGCCCGCGGTGCATGTGCTGCACCTGAGC 5220  
Qy 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
Db 5221 AACATTAACGTGGGGGAATACCTGGGCCAGCGCGGGGTGTGGCCCTTGTGGGGC 5280  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGlnGlyValaSerSerLeuAspProSer 1780  
Db 5281 TGTTTGACGGGTGTGGGTGATGCGTACGCGCAGGGGGTTAACGCTGGATCCAGC 5340  
Qy 1781 HisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGACATTCMAGTGAGCAGAGCTGTACCTGTGACCTTGTGACCTCAAC 5400  
Qy 1801 ProCysProAlaAsnSerThrCysSerAsnAspThrAspSerThrCysCysSerCysAsp 1820  
Db 5401 CGGTGTCTGCTTAACGCTATTGTGACGACACAGTGGGACAGCTATTCTGTGAGCTGTAT 5460

QY 1821 ProGlyTyrTyrGlyAspAsnGlyThrAsnValCysAspLeuAsnProGlyLysGln 1840  
 Db 5461 CCAAGTTACTATGATGTGACAACTGATATGATGTGACCTGAACCCGTTGTAGACACAG 5520  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrThrCysGlyCysProProAsn 1860  
 Db 5521 TCTGTGTGTATCCCGCAAGCCCAAGTGTGCCCCCATAGGCTATACCTGCGAGTGTCCCAAAAT 5580  
 QY 1861 TyrLeuGlyProTyrCysGlyLysThrArgLysAspGlyProCysProArgGlyTyrPrtPrtGly 1880  
 Db 5581 TACCTTGAGCCATACCTGTGAGAACCGAGATTGACAGCCTTGTCCTCCGCTGTGTGGGA 5640  
 QY 1881 HisProThrCysGlyProCysAsnGlyAspValSerLysGlyPheAspProAspCysAsn 1900  
 Db 5641 CATCCCAACATGTGGCCCATGCACTGTGATGTGCAAGAAAGCTTTGACCCAGACTGCAAC 5700  
 QY 1901 LysThrSerGlyGlyCysHisGlyLysGlyLysAsnHisGlyTyrArgProProGlySerProThr 1920  
 Db 5701 AAGACAAAGCCGAGGTGCCACTGCAAGAGAACCACTAACCGGCCCCAGAGCAAGCCACCC 5760  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyLysAsp 1940  
 Db 5761 TGCCTCTGTGTGATGCTACCCCAAGGCTCTTGTCAGAGCTGTGACCTGTAGAGAT 5820  
 QY 1941 GlyGlnCysProCysLysProGlyValLysGlyArgGlnCysAspArgCysAspAsnPro 1960  
 Db 5821 GGCACAGTGTCAATGACCAAGGAGTGTGAGGAGTGTGATGATGATGATGATGATGATGATGAT 5880  
 QY 1961 PheAlaGlyValThrThrAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaLys 1980  
 Db 5881 TTGTGAGAGGTGACCAACCAATGTGCTGTGAAGTATATGACAGCTGCCCAACAGAGGAT 5940  
 QY 1981 GluAlaGlyLysTyrPrtProArgThrArgPheGlyLeuProAlaAlaAspCysPro 2000  
 Db 5941 GAGGCTGGAGATCTGTGTGCCCCGATCCCGCTGAGGCTGTGCTGTGCTGTGCTGTGCTGTGCT 6000  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlyLysAspGlyTyrLeuProPro 2020  
 Db 6001 AAAGGCTCTTTGGAGCTGCTGTGCGGCACTGTGATGAGACAGGGGGGTGTGCTCCCCCA 6060  
 QY 2021 AsnLeuPheAsnGlyThrSerLysThrPheSerGlyLeuLysGlyPheAlaGlyLysLeu 2040  
 Db 6061 AACCTCTTCAACATGACGTCATCACTTCTCAGAACTGAAGGGCTTGTGTGAGCGGCTA 6120  
 QY 2041 GlnArgAsnGlyLysSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
 Db 6121 CACCGAATATGATCAAGGCTTGAATCAAGGCGCTCCAGACGCTAGCCCTGTCTGTGCTG 6180  
 QY 2061 AsnAlaThrGlnHisGlyThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
 Db 6181 AAGCGCACGACAGACACAGCTGCTACTTGGCAGACGACGTCAGAGGTGGCTTACAGCTG 6240  
 QY 2081 AlaThrArgLeuLeuAlaHisGlyLysThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 Db 6241 GCCACCGGCTGTGTGCGCCACAGAGACACCAAGGGGCTTGTGTGTGTGTGTGTGTGTGTGT 6300  
 QY 2101 AspValHisPheThrGlnLeuLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
 Db 6301 GAGCTGACCTTCACTGAGATCTGCTGTGCGGTGTGCGAGGCGCTTCTGTGACACAGCCAC 6360  
 QY 2121 LysArgHisGlyTyrGlyLeuLysGlnGlnThrGlnGlyGlyTyrAlaTyrLeuLeuGlnHis 2140  
 Db 6361 AAGCGGACCTGGAGCTGATCCAGACAGACAGAGGTGTGACCGCTGTGTGTGTGTGTGTGTGT 6420  
 QY 2141 TyrGlnAlaLysTyrAlaSerAlaLeuAlaGlnAsnMetArgHisGlyTyrLeuSerProPhe 2160  
 Db 6421 TATGAGGCTTACCGCAGGT 6480  
 QY 2161 ThrIleValThrProAsnIleValLysSerValValArgLeuAspLysGlyAsnPheAla 2180  
 Db 6481 ACCATGTGTACGCGCCACATTTGTATCTCCGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6540

QY 2181 GlyAlaLysLeuProArgTyrGlyLysLeuArgGlyGlnGlnProProAspLeuGlyLys 2200  
 Db 6541 GGGGGCCAACTCTCCCGCTACAGAGGCGCTGTGGGAGACACCCCGGACCTTGTAGACA 6600  
 QY 2201 ThrValIleLeuProGlyLysValPheArgGlyLysThrProValValArgProAlaGly 2220  
 Db 6601 ACAGTCATCTCTGAGTGTGTCTTTCAGAGAGAGAGCCCGGTGTGTGTGTGTGTGTGTGTGT 6660  
 QY 2221 ProGlyLysAlaGlyLysLeuProGlyLysLeuAlaArgArgGlyArgArgHisAspProLysLeu 2240  
 Db 6661 CCGGAGAGGCGCCAGAGGCGACAGAGCTGTGCAAGGCGACAGAGGCGACCCGAGAGCTG 6720  
 QY 2241 SerGlnGlyLysAlaValAlaSerValIleLysTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 6721 AGCAGAGGAGGCTGTGTGCGCAGCGGTATCATCTACCGCACCTGTGGCGGGCTGTAGCTCT 6780  
 QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleLysAsnThr 2280  
 Db 6781 CATTAATATGACCTGTGACAAAGCGCAGCTTGAGAGTGTCCCAACCGCCGATCATCAACA 6840  
 QY 2281 ProValIleSerLysSerValHisAspAspGlyLysLeuLeuProArgAlaLeuAspLys 2300  
 Db 6841 CCGGTGTGAGCATACGCTCATATGATGAGAGCTTCTGCCCCGGGCGCTGTGACAA 6900  
 QY 2301 ProValThrValGlnPheArgLeuLeuGlyLysThrLysProIleCysVal 2320  
 Db 6901 CCGGTACAGGTCAAGTTCGCGCTGTGTGAGACAGAGGAGCGGACCAAGCCCATCTGTGTG 6960  
 QY 2321 PheTyrAsnHisSerLysLeuValSerGlyTyrThrGlyTyrPrtSerAlaArgGlyCysGly 2340  
 Db 6961 TTCTGTGACATTCATCACTGT 7020  
 QY 2341 ValIlePheArgAsnGlyLysSerHisValSerCysGlnCysAsnHisGlyMetThrSerPheAla 2360  
 Db 7021 GTGCTTCTCCGAAATGAGAGCCACTGACCTGCGAGTGTGACACCACTGACAGCTTCGCT 7080  
 QY 2361 ValLeuMetAspValSerArgArgLysGlnGlyLysLeuProLeuLysThrLeuThr 2380  
 Db 7081 GTGTGTATGACGTTTCTGT 7140  
 QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400  
 Db 7141 TACGT 7200  
 QY 2401 LeuArgIleLeuArgSerAsnGlnHisGlyLysArgArgAsnLeuThrAlaAlaLeuGly 2420  
 Db 7201 TTGCGATCTCTGCTGT 7260  
 QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyLysLeuGlnAlaAspLeuProPheAlaCysThr 2440  
 Db 7261 CTGGCTCAGCTGTGTCTTCTCTGT 7320  
 QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGly 2460  
 Db 7321 GTCATTTGCCATCTGT 7380  
 QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValAlaAspValAsnThrGlyProMetArg 2480  
 Db 7381 GCTTGTACCTTGTACCGGCACTGT 7440  
 QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
 Db 7441 TTCTACTATCTGT 7500  
 QY 2501 AspProGlyLysTyrGlyLysAspProAspPheCysTyrPheLeuSerLysTyrAspThrLeuIle 2520  
 Db 7501 GACCCGAGAGGCTTACGGGAACTGT 7560  
 QY 2521 TyrPhePheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
 Db 7561 TGGAGTTTGT 7620  
 QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyLysGlyProValSer 2560



Db 7621 GCGGCGCCGCTCTGTCGTCCTCCAGCGGCGCTTGAAGAAAGTCTCTCTCG 7680  
 Oy 2561 G1yleuGlnProSerPheAlaValleuLeuLeuSerAlaThrThrPheLeuAlaLeu 2580  
 Db 7681 GGCCTGAGAGCCCTCTGCGCGCTCTCTGCTGCTGAGGCCACGTGGCTGTGACATG 7740  
 Oy 2581 LeuSerValAenSerAenThrLeuLeuPheHisTyrLeuPheAlaThrCysAenCysIle 2600  
 Db 7741 CTCTGTGTCACAGGACACCTCTCTCTGCTCACTACTTGTGCTACCTGCAATGTCATC 7800  
 Oy 2601 G1ng1yProPhe11PheLeuSerTyrVal11LeuSerTyrVal11ArgTyrVal11Leu 2620  
 Db 7801 CAGGCGCCCTTATCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7860  
 Oy 2621 LysLeuAlaCysSerArgTyrSerProSerProAlaLeuThrThrLeuSerThrLeu 2640  
 Db 7861 AAGCTTGCTGAGCGCGAGCCAGCCAGCCCTGACCTGCTGCTGCTGCTGCTGCTGCTG 7920  
 Oy 2641 ThrSerSerTyrAenCysProSerProTyrAlaAenArgTyrLeuTyrGlnProTyrGly 2660  
 Db 7921 ACCTGCTCTTACACTGCGCCAGCCCTTACAGAGTGGCGGCTGCTGCTGCTGCTGCTG 7980  
 Oy 2661 AenSerAlaGlySerLeuHisSerThrSerArgSerGlyTyrSerGlnProSerTyrIle 2680  
 Db 7981 GACTGCGCGGCTCTCTGACAGCACCACTGCTGCGGCGAGAGTCAAGCCAGCTACATC 8040  
 Oy 2681 ProPheLeuLeuArgTyrGlnuSerAlaLeuAenProGlyGlnGlyProProGlyLeuGly 2700  
 Db 8041 CCTTCTTCTGAGGAGAGAGTCCGCACTGAACTGCGCAAGGCCCTTGGCTGGGG 8100  
 Oy 2701 AenProGlySerLeuPheLeuGlnGlyGlnAenGlnHisAenProAenThrAenSer 2720  
 Db 8101 GATCCAGGAGCGCTCTCTCTGAGAGGTCAAGACAGCAGCATGATCTGACAGGACTCC 8160  
 Oy 2721 AenSerAenLeuSerLeuGlnuAenAenGlnuSerTyrAlaSerThrHisSerSer 2740  
 Db 8161 GACACTGACCTCTCTTGAAGACACCAAGATGCTCTTGAAGCTTCAACCACTACATCA 8220  
 Oy 2741 AenSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 Db 8221 GACACTGAGGAG 8280  
 Oy 2761 TrpAenSerLeuLeuGlyProGlyAlaGlnArgLeuProLeuHisSerThrProLysAen 2780  
 Db 8281 TGGGATAGCTGCTGCGGCTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8340  
 Oy 2781 G1yG1yProGlyProGlyTyrAlaProTyrProGlyAenPheGlyThrThrAlaLeuGlu 2800  
 Db 8341 GGGGCGCCAGGCGCTGAG 8400  
 Oy 2801 SerSerGlyAenGlyAlaProGlnGlnArgLeuArgTyrGlnAenGlyAenAlaLeuSerArg 2820  
 Db 8401 AGTACTGGGAG 8460  
 Oy 2821 G1ng1ySerLeuGlyProLeuProGlySerSerAlaGlnProHisLeuGly 2837  
 Db 8461 GAGGCGCTCTGAGCGCCCTTCAAGGCTCTTCTGCGCAGCTCACAAAGGT 8511

RESULT 5  
 Db 76469 D87469 8924 bp mRNA linear PRI 06-OCT-2001  
 LOCUS Human mRNA for KIAA0279 gene, partial cds.  
 DEFINITION D87469  
 ACCESSION D87469.1 GI:1665820  
 VERSION D87469.1  
 KEYWORDS KIAA0279.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y.,  
 Chara, O., Tanaka, A., Kotani, H., Miyajima, N. and Nomura, N.

TITLE Prediction of the coding sequences of unidentified human genes. VI.  
 The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 analysis of cDNA clones from cell line KG-1 and brain  
 JOURNAL DNA Res. 3 (5), 321-329 (1996)  
 MEDLINE 97191544  
 PubMed 9039502  
 REFERENCE 2 (bases 1 to 8924)  
 AUTHORS Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-AUG-1996) Otsu Ohara, Kazusa DNA Research Institute;  
 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)  
 FEATURES  
 Location/Qualifiers  
 source 1..8924  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HA6133"  
 /sex="male"  
 /issue\_type="brain"  
 /clone\_id="pSPORT 1"  
 1..8924  
 /gene="KIAA0279"  
 <1..7227  
 /gene="KIAA0279"  
 /note="Similar to D.melanogaster cadherin-related tumor  
 suppressor"  
 /codon\_start=1  
 /protein\_id="PAA13407.1"  
 /db\_xref="GI:1665821"  
 /translation="ESVPLGSLVLAVALADADADNARLEAYLAGVGHDPPTINCT  
 GMLSVALELREEDVPSFGVBARHDTGTPALTSASVSVVLVDVNDNPPETDPEYTV  
 RLNEDAVGVSVVAVSDADHSHVITQITSGTNRPSITSGGGLVSLAPLDY  
 KLEQVYLAATASGTRDPAQIVVNTQDANTHRPVOSHSHTVNVNEDRPAGTVVL  
 ISATDPTGEMARITVPMDSIIPQRIDATGATVDAEIDYEDVQSVTALATARDG  
 IPQSGDTTYEILVNDVNDNAPRDRPSGSGVYEDVPPPTSVQIATDRDSELNR  
 VPTFGGDDGDDGVFESTSGIVTRLRRLDRENAVQVLAIVADVGMPARPMVEY  
 TVVVLVDNDNPVPEODEPVPVENSPVGLAVARVATPDDETNQIQIVQEGNI  
 PEVFDLDIFSGELVALVDLDYBDRPEVLAVALVATVATVNLNDNPVYL  
 GNPEILFNNTVYNNSSSPGGAIGRVAPHPDIDSLTYSERENELSLVLAASGE  
 LKLSRALDNNRPLEAIVSVLSDGVSHVTQAQIVITIDEMLTHTSITLRLDMSPE  
 RLSPILGLPITQAVATLATLPDHPVIVPNQRTDADAGHILNVSLVCGPRGCGGP  
 PFLPSDELQERLYLNRSLTLAISQRLVPPDNLCLPEPCNTRKCVSRSEGGAP  
 IASSSVLEPRLHPVGLRCRCRPGEDVCEDEVLDLCEYSRCPGSHGCRSSEGGATCL  
 CRDYGTEHGVSNARSGRCTPGVCKNGCTCNILVGGFKDCPSGDEPKCYQVTTTS  
 PPASFTPTFGLRFRFHTLALSPATKEROCLLNGRFNKGFPVLALEVIQEOVLT  
 PSAGESTTVSPVPGVSDGOMHTVOLKYNNELDQTLPOGSRGKVAVTVDDC  
 DTGVALRGSVLYNYSCAAGTGGSKSLDTLPPLDGLGTPDLPSFPVRMRPVC  
 MRNLQVDSRHLDMADFLANNGTVPGCPAKKNVCDNCTCHNGTGVNDAPFSCPLG  
 FGKSCAQEMANPQHFAGSLVAMHGLSLPISQWYLSMERTQADGVLLQALTRGR  
 STITLQLRBGMVMSVEGTGLQASLSLBRANDGMHQAOLAGSGGHAITLRF  
 DYGOQRALNGLPGLHLHLNLTIVGIPBPAAGVAFGRCLGAGVSDPEVNSI  
 DPHSGSINVEGCSLPDPCDSNCPGAGVSNMDVSGSCDGGYGDNTCNIDL  
 PCRHSQCTKRPSPHGYTCPCPNYIAPYCETRIIDPCRGMGHPTCCGCNDYVK  
 GFDDPNKTGSECHCKENHYRPPESPTCLLDCTPTGSLSLVCDPEGQCEKRGVIG  
 RQDCRCNPFAEVTCNCEVNVDSCPAIFAGIWPPTRLPAPACPKPSKPTAVR  
 HCDHRGMFLPNLNFCTSIYTSBKGAELQORNSGLDGRSCQALLLNAQOHA  
 GFSGSDVAVYOLATRLRLHSTORGGLNATQDVTENLTARGLALLDTPANRHE  
 LIQTBGTAMLIQHYBAVASALANMRHYLSPTITVNTYIVSVRLRQNRHPS  
 LPRYBARGRQPPHETTVLILPESVPEPTPVYRPAPEGARQEBELRQRHPS  
 QGRVAVVITRTLAGLPHNYDDPKRSILVPRKPIINTPVSLSVHDEBLRPLAD  
 KPVTVQRLRLREBERTPILCPMNHSLIVGCTGMSRGEVFRNRSVSCQCNHT  
 SPVATLNDVSRREBERILPLKTLVYALGVTLAALLPFLTLRLIRSNHGIRNL  
 TAAQLAQVLQINQADLPACTVAILIHLFLCTPSWALTEALHLYALTEVAD  
 VNTGPMRFYMLGNGVPAFTGLAVGDPGNGPDCMWSIYDTLMSFPQVPAFV  
 SMSVFLYTLAARSCAAROGFRKKGVSGLQSPAVLLLSATWLLLSVNDTL  
 FHTIPATCKTQGPFTFLSYVLSKVRKALKLACSKRPSPDPAITLTKSTITSNCP  
 SPVADGRLYQPYGDSAGLSHSTRSGSQSYVLPFLRESBALPQGPGLGDPGL  
 FLGQADRDHPDSDSLBLBDQSGSVASTHSSDEEBEEREAAPFGEGWMS  
 LLEGQARPLHSPKDGCGPKAPWPGTGTAKSSGGAPEERLRNGDLSSE  
 GSLGFLPGSSAOPHKGLIKKCLPTIKSLRLPLEOCSSGSRGSSASBGSGGPP  
 PRPPEPRLDEOLNGWPIRIMSIKAGVVDSDSSSERLFFNFPLH"

BASE COUNT 1727 a 2846 c 2504 g 1847 t

## ORIGIN

## Alignment Scores:

Pred. No.: 0  
 Score: 2408.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 82.38%

DB: 9

Length: 8924  
 Matches: 2408  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-916-849A-3 (1-2923) x DB7469 (1-8924)

QY 516 GluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGly 535  
 DB 1 GAAGAGGTCCTCCCTTAAGCTACCTGGTCTCCATGTCACAGGCTTACACGCTGATGCTGGT 60  
 QY 536 AspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIle 555  
 DB 61 GACAAATGCCCGCTGGAAATACCCCTTGCCTGGGGTGGACATGATCTCCCTTACCAATC 120  
 QY 556 AsnAsnGlyTyrGlyTyrPheSerValAlaAlaGlyLeuAspArgGluGluValAspPhe 575  
 DB 121 AACAAATGGCACAGGCTGATCTCTGGTGGCTGAAGCTGGACCGGAGGAAGTTGATTC 180  
 QY 576 TyrSerPheGlyValGlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSer 595  
 DB 181 TACAGCTTGGGGGTAGAAAGCTCAGACCAATGAGCACTCCAGCACTCATGCTGGCCAAAT 240  
 QY 596 ValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnProGluTyr 615  
 DB 241 GTACAGCTGATCTCTCTGATGTCACAGCAACAACTTTCACCAACCAAGATTAC 300  
 QY 616 ThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaVal 635  
 DB 301 ACAGTCGGGCTCAATAGAGATGACAGCTGTGGCACAGCGTGTGACGGTGTCAAGCTGTG 360  
 QY 636 AspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArg 655  
 DB 361 GACCGGATGCTCATAGTGCATACCTTACCAATGACCAATGACCAATGCTCAAAACCCG 420  
 QY 656 PheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyr 675  
 DB 421 TTCTCATCAACCAAGCAAGTGGTGGCTGTATCCCTTCCCTGCACTGACATAC 480  
 QY 676 LysLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThr 695  
 DB 481 AAACCTGAGCGGCAATGTGTGGCTGTACCGCTCCGATGCACTCGGCGAGCAACG 540  
 QY 696 AlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSer 715  
 DB 541 GCAACAGATGTGTGAATGTACCGACGCCAACACCCATGCTCTGTCTTTCAGAGCTCC 600  
 QY 716 HisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSer 735  
 DB 601 CACTATACAGTGAATGTTATATGAGAACCGCGCGGAGGACCAAGGGTGGCTGATCAG 660  
 QY 736 AlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIle 755  
 DB 661 GCCACGATGAGGACACAGGTGGAATGCGCGCATCACTTACCTTCAATGGAGGACACATC 720  
 QY 756 ProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlyLeuAspTyr 775  
 DB 721 CCCACGTTCCGATCAGATGACAGACCGGGGGCTGTACCAACCGGTGAGCTGAGACTAC 780  
 QY 776 GluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLys 795  
 DB 781 GAAGACCAAGTCTTACACCCCTGGCCATTACTGCTGGGCAATAGGCATTTCCCGAAG 840  
 QY 796 SerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPhe 815  
 DB 841 TCCGACCAACCACTTACCGGATCTGGTGAACGACCTGATATGACATATGCCCTCAGTTT 900  
 QY 816 LeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerValLeu 835

DB 901 CTGCAAGACTCTTACCAAGGAGGTCTATAGAGATGAGCCACCTTCATTAAGCTCTCG 960  
 QY 836 GlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGln 855  
 DB 961 CAGATCTCAGGCCACCTGATCGTATCTTGACATTAATGCAAGGGTCTTTCACCTTCCA 1020  
 QY 856 GlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThr 875  
 DB 1021 GGAGGCGACGATGAGACGGTGAATTTATGTGTGATGCCAGTCAGGACATCGTGGAAAG 1080  
 QY 876 LeuArgValLeuAspArgGluAsnValAlaGlnTyrValIleAspArgAlaTyrAlaValAsp 895  
 DB 1081 CTACGAGAGCTGATGAGAGAACCTGGCCAGATATGCTTGGCGGGCAATAGCAATGAGAC 1140  
 QY 896 LysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAspValAsn 915  
 DB 1141 AAGGGAGATGCCCGACCGCCGACACTTATGGAAGTACAGTCACTGTGTGGATGATGAT 1200  
 QY 916 AspAsnProProValPheGluGlnAspGluAspValPheValGluGluAsnSerPro 935  
 DB 1201 GACAAATCCCGCTGCTTGAAGCAGGATGATTTGATGTGTGTGGAAGAGAACAGCCCC 1260  
 QY 936 IleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAspAlaGln 955  
 DB 1261 ATTTGGGCTAGCCGTGGCCCGGGGTCAACGCCACTGACCCCGATGAAAGCAACATGCCAG 1320  
 QY 956 IleMetTyrGlnIleValAlaGlyValAsnIleProGluValPheGlnLeuAspIlePheSer 975  
 DB 1321 ATTATGTACAGATTTGTGAGGGGCAACATCCCTGATAGCTTTCAGAGTGGACATCTTCC 1380  
 QY 976 GlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuVal 995  
 DB 1381 GGGGAGCTGACAGCCCTGTGTAGACTTAACAGAGACCGCGCTGATAGTCTCTGCTTC 1440  
 QY 996 IleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAsp 1015  
 DB 1441 ATCCAGGCCACAGTCAAGCTCTCTGTGTAGCCGGGTACAGTCAAGTCCGCTCTTAC 1500  
 QY 1016 ArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThr 1035  
 DB 1501 CGCAATGACAAACCAACCAAGCTGTGGCAACTTGTAGATCTTTTCAACATATGTACAC 1560  
 QY 1036 AsnArgSerSerSerPheProGlyGlyValAlaIleGlyArgValProAlaHisAspProAsp 1055  
 DB 1561 AATCCCTCAAGAGAGCTTCCCTGGGGGTGCCATTTGCCAATGACCCCATGACCTTAT 1620  
 QY 1056 IleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuLeu 1075  
 DB 1621 ATCTCAGATAGTCTACCTTACAGCTTTAGCGGGAAATGAATCTAGGCTGTGCTCTC 1680  
 QY 1076 AsnAlaSerThrGlyGlyLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlu 1095  
 DB 1681 AATGCTTCCACGGGGTGAAGCTTAAGCCCGGACCTGACCAACACCGGCTCTGGAG 1740  
 QY 1096 AlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGlnCysValLeu 1115  
 DB 1741 GCCATCAATAGAGGTCTGTGTGACAGCGCGTACACAGCGTGAACCGCCAGTGGCCGCTG 1800  
 QY 1116 ArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAsp 1135  
 DB 1801 CGTGTACATATATACCAAGATGATGCTTACCCACAGATACAGCTGCGCTGGAGAGAC 1860  
 QY 1136 MetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAla 1155  
 DB 1861 ATGTCAACCGAGCGCTTCTGTCAACCACTGAGGCTCTTCACTCAAGCGGTGGCCGCC 1920  
 QY 1156 ThrLeuAlaThrProProAspHisValValValPheAsnValGlnArgAspThrAspAla 1175  
 DB 1921 ACGTGGCCACCGCACCGGACCACTGTGTGTCTTCAACGTAACAGCGGAGACCCAGCGCC 1980  
 QY 1176 ProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProGlyProGlyGly 1195



Dh 1981 CCGGGGGCCACATCTCAACGTGAGCCTGTGGTGGGCCACGCCGCGGAGGCGGCGGCGC 2040  
Qy 1196 G1yProBpPheLeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeu 1215  
Db 2041 GGGCGCGCCCTTCCGCTCGAGGACCTGAGAGAGGCGCTATATCTCAACCGGAGCGTGG 2100  
Qy 1216 LeuThrAlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCySLeuArgGlu 1235  
Db 2101 CTAGCGGCACATCTCGGACAGCGCGTGTGCTCTTCCACAGCAACATCTCTGCGGAGG 2160  
Qy 1236 ProCysGluAsnTyrMetArgCysValSerValLeuAspPheAspSerSerAlaProPhe 1255  
Db 2161 CCTCGAGAACTCAATGCGCTGCGTGGTGGCTGCTCACTCTCGGCGCTTC 2220  
Qy 1256 IleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArg 1275  
Db 2221 ATGCGCTCCTCCTCGGCTCTTCCGCGCATCACCCGTCGAGGGCTGCGCTGCGCGC 2280  
Qy 1276 CysProBpGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCySlyrSerArg 1295  
Db 2281 TGGCGCGCCGCTTCACGGGTGACTACTGCGAGACCGAGGTGGACCTCTGCTCGCGG 2340  
Qy 1296 ProCysGlyProHisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCySArg 1315  
Db 2341 CCTGTGCCCCCAAGGGCGCTGCGGACGCGGAGGGCGGCTACACTGCTCTGTCTGT 2400  
Qy 1316 AspGlyTyrThrGlyGlnHisCysGluValSerAlaArgSerGlyArgCysThrProGly 1335  
Db 2401 GATGGCTACACGGGTGAGCACTGTGAGGTAGTGTGCTCAAGCGCTTGCACCCCGGT 2460  
Qy 1336 ValCysLeuAsnGlyGlyTyrCysValAsnLeuLeuValGlyGlyPheLysCysAspCys 1355  
Db 2461 GTCTGCAAGATGAGGGCACCTGTGTCAACTGTGTGGCGGGTTCAAGTCCGATTGC 2520  
Qy 1356 ProSerGlyAspPheGlyAspProLysProLysCysGlnValThrThrArgSerPheProAlaHis 1375  
Db 2521 CCACTGAGAACTTCAGAAAGCCCTACTGCGAGGTGACACGCGAGCTTCCCGGCCAC 2580  
Qy 1376 SerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPhe 1395  
Db 2581 TCTTATCATCACTTTCGGCGCGCTGCGGCAACGTTTCCACTTCACTCGGCGCTCTGT 2640  
Qy 1396 AlaThrLysGlyLysArgGlyLeuLeuLeuTyrAsnGlyValArgPheAsnGlyLysHisAsp 1415  
Db 2641 GCCACAAAGAGGCGCACGGGTGTGCTGTGTACAAATGGCGTTCAATGAGGAAGCATGAC 2700  
Qy 1416 PheValAlaLeuGlnValIleGlnGlnGlnValAlaGlnLeuThrPheSerAlaGlyGlyLeuSer 1435  
Db 2701 TTGTGGCCCTCGAGGTGATCCAGAGCAAGGTCAAGCTCACTTCTGTGAGGGGAGTCA 2760  
Qy 1436 ThrThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrVal 1455  
Db 2761 ACCACCAAGGTGCTCCCATTTCCGTGCGCGGAGGTGAGTGTGCGAGTGGCGATGACGTG 2820  
Qy 1456 GlnLeuLysTyrTyrAsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSer 1475  
Db 2821 CAGCTAAATATCAATACATTAAGCACTGTGGGTCAACAAGGGCTCCCAAGGCGCATCA 2880  
Qy 1476 GlnGlnLysValAlaValValThrValAspGlyCysAspThrGlyValAlaLeuAspPhe 1495  
Db 2881 GAGCAGAAAGGTGCTGTGTGACCGTGTGATGCGTGTGACACAGAAAGTGGCTTGGCCTTC 2940  
Qy 1496 GlySerValLeuGlyAsnTyrSerCysAlaAlaGlnGlyTyrGlnGlyGlySerLysLys 1515  
Db 2941 GAGTCTGTCTGGGCAACTACTCTGTGTGCTGCGCCAGGGCACCCAGGTGGCAGCAAGAG 3000  
Qy 1516 SerLeuAspLeuThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPhe 1535  
Db 3001 TCTCTGATCTGACGGGGCCCTGTCTACTAGCGGGGTGCTGACCTGCGCGAGACCTTC 3060  
Qy 1536 ProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHis 1555  
Db 3061 CCAAGTCCGAATGGGAGATTCTGTGGCTGACATCGGAACCTTGAGGTGACAGCCGGAGC 3120

Qy 1556 IleAspMetAlaAspPheIleAlaAsnGlnTyrThrValProGlyCysProAlaLysLys 1575  
Db 3121 ATAGACATGGCTGACTTCACTTTCGCAACATGGCACCTGTGCTGCGCTGCGCGCAAGAG 3180  
Qy 1576 AsnValCysAspSerAsnThrCysHisAsnGlyGlyTyrCysValAsnGlnThrAspAla 1595  
Db 3181 AACGTGTGTGACAGCAACTTGGCCAAATGGGGGGCACTTGGCGTGAACAGTGGAGCGCG 3240  
Qy 1596 PheSerCysGlyCysProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsn 1615  
Db 3241 TTGAGCTGAGAGTCCCCCTGGAGCTTGGAGGCGAAGCTGGCGCCAGAAATGGCCAAAT 3300  
Qy 1616 ProGlnHisPheLeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSer 1635  
Db 3301 CCAAGACACTTCTCGGAGCAGAGCCTGTGCTGCGCAATGGCTCTCGCTGCGCACTTC 3360  
Qy 1636 GlnProTyrTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGln 1655  
Db 3361 CAACCTGGTACCTCAGCTCATGTTCGCGACGCGCAGCGCCGACGGTGTCTGTGACAG 3420  
Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisIleValMetLeu 1675  
Db 3421 GCCATCACAGGAGGGCGGACAGCAATCACTTACAGCTACAGAGGGCCACGTATGTG 3480  
Qy 1676 SerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyLysArgAlaAsn 1695  
Db 3481 AGCGTGAAGGGACAGGGCTTACAGGCTCTCTCTCCGCTGGAAGCAGGCGGCGCAAT 3540  
Qy 1696 AspGlyAspThrPheHisAlaGlnLeuAlaLeuGlyValAspGlyGlyProGlyHisAla 1715  
Db 3541 GACGGTGACTGACCAATGACAGCTGGGCACTGGAGCGACAGGGGGGGCGCTGGCCATGCC 3600  
Qy 1716 IleLeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValAsnLeuGlyProArgLeuHis 1735  
Db 3601 ATTCTGCTCTTGATTAATGAGGAGGACAGAGAGGAGGAGAACCTGGAGCCCGCGCTGACAT 3660  
Qy 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
Db 3661 GGTCTGACCTTACGACATTAACATGAGGGCGAAATCTCGGGCCAGCGCGGCTGTGGCGC 3720  
Qy 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsn 1775  
Db 3721 CGTGGCTTTCGGGGCTGTGTGAGGGGTGTGCGGGTGAAGCAATCCAGAGGGGGTTAAC 3780  
Qy 1776 SerLeuAspProSerHisGlyGlyLeuSerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
Db 3781 AGCTGTGATCCAGCCATGAGGAGAGAGCATCAACGTGAGCAAGGCTGTAGCCTGCTGAC 3840  
Qy 1796 ProCysAspSerAspProCysProAlaAsnSerTyrCysSerAspAspThrAspSerTyr 1815  
Db 3841 CTTTGTGACTCAACCGGTCTCTGCTTAACAGCTATTGAGACAGACCTGGGACAGCTAT 3900  
Qy 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
Db 3901 TCTTGACGTGTGATCCAGGTACTATGTGTGCAACTGACTAAATGTGTGTGACCTGAAAC 3960  
Qy 1836 ProCysGlnHisGlnSerValCysThrArgLysProSerAlaProHisGlyTyrThrCys 1855  
Db 3961 CCGTGTGACACAGTGTGTGTATCCGCAAGGCCAGAGGCCCCCAATGGCTATACCTGAC 4020  
Qy 1856 GlnCysProProAsnTyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
Db 4021 GAGTGTCCCCCAAAATTAACCTTGGGCAATACGTGTGACCAAGGATTGACCAAGCTTGTCCC 4080  
Qy 1876 ArgGlyTyrThrPheGlyHisProThrCysGlyProCysAsnCysAspValSerLysGlyPhe 1895  
Db 4081 CGTGTGTGTGGGCAATCCCAATGTGCGCCCAATGACCTGATGATGACGAAAGGCTTT 4140  
Qy 1896 AspProAspCysAsnLysThrSerGlyGlyCysHisCysLysGlnAsnHisIleTyrArgPro 1915  
Db 4141 GACCGAGACTGACAAAGCAAGAGCGGCGAGTGTCCACTGAGAGAAACCACTACCGGCGCC 4200

QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 DB 4201 CAGAGCAGCCCACTGCTCTTGTGTGATGCTGCAACCCCAAGGCTCTTGTCCAGAGTC 4260  
 QY 1936 CysAspProGlyLysArgGlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAsp 1955  
 DB 4261 TGTGACCTCAGAGATGCGCAGTGTCCATGCAAGCCAGGTGTCACTCGGCGCTCAGTGTGAC 4320  
 QY 1956 ArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSer 1975  
 DB 4321 CGCTGTACAACTTTGCTGAGGTCAACAATGCTGTGAAAGTGAATATGAGACG 4380  
 QY 1976 CysProArgAlaIleGluValGlyIleTyrProArgThrArgPheGlyLeuProAla 1995  
 DB 4381 TGCCCAAGAGGATGAGGCTGGGATCTGGTGGCCCGGTACCGGCTTGCGGCTGCT 4440  
 QY 1996 AlaAlaProCysProLysGlySerPheGlyThrAlaValArgIleCysAspGlyLysArg 2015  
 DB 4441 GCTGCTCCCTGTCCCAAGGCTCCTTGGGACTGCTGTGGCCACTGTGATGAGCAAGG 4500  
 QY 2016 GlyTyrLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLysGly 2035  
 DB 4501 GGGTGGCTCCCCCAAACTCTTCACTGACGTCATCACTTCTCAGAACTGAAAGGAC 4560  
 QY 2036 PheAlaGluArgLeuGlnArgAsnGlyLysSerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 DB 4561 TTGCGTAGAGGGCTACAGCGGAATGAGTCAAGGCTTAACTCAAGGCGCTCCCAAGAGCTA 4620  
 QY 2056 AlaLeuLeuLeuLysArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLys 2075  
 DB 4621 GCCCTGCTCTGGCCAGCAGCGCAAGCAGCAAGCTGTCTACTTCGCGACGACGTCACAG 4680  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlyLysSerThrGlnArgGlyLysPheGly 2095  
 DB 4681 GTGGCTCAACAGCTGGCCAGCGGCTCTGGCCCAAGAGCAAGCCAGCGGCTTTGGG 4740  
 DB 4741 CTGTCTGCCACACAGAGCGTGCACCTTCACTGAGAAATCTGCTGGCGGGGCAAGGCGCTTC 4800  
 QY 2116 LeuAspThrAlaAsnLysArgGlyIleTyrGlnLeuLleGlnGlnThrGlyLysGlyTyrAla 2135  
 DB 4801 CTGACACAGCCCAACAGCGGCACTGGAGCTGATTCAGACAGAGAGGGTGGACCGCC 4860  
 QY 2136 TyrLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetCysArgIleThr 2155  
 DB 4861 TGCGTGTCCAGCACTATGAGGCTTACGCAAGTGCCTTGGCCCAAGAACATGGCGCACACC 4920  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
 DB 4921 TACCTAAGCCCTTCACTCATCTGTCAAGCCCAATGTGTCACTCCGTAGTGGCGCTTGAC 4980  
 QY 2176 LysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGlnAlaLeuLysGlyLysGlnPro 2195  
 DB 4981 AAAGGGAACTTGTGGGGCCAGAGTGCCTCCGCTACAGAGCCCTGCTGGGAGAGCAGCCC 5040  
 QY 2196 ProAspLeuGlnThrThrValIleLeuProGlyLysSerValPheArgGlyLysThrProProVal 2215  
 DB 5041 CCGGACCTTGAAGACAAGTCACTTCTGCTGAGTCTGTCTTCAAGAGAGAGCCCGCCG 5100  
 QY 2216 ValArgProAlaGlyLysGlyLysGlnGlnGlnProGlyLysLeuAlaArgArgGlnArg 2235  
 DB 5101 GTACAGGCGCCAGAGCCCGGAGAGGCCCAAGAGCCAGAGAGGCTGGCAAGCGGCAAGCGA 5160  
 QY 2236 ArgHisProGlyLeuSerGlnGlyLysValAlaSerValIleIleTyrArgThrLeu 2255  
 DB 5161 CGGAGCCCGAGGCTGAGCCAGGGTGAAGCTGTGGCCAGCTCACTCACTCAAGACCTG 5220  
 QY 2256 AlaGlyLeuLeuProHisAsnTyrAspProAspLysArgSerLeuArgValProLysArg 2275  
 DB 5221 GCGGGGCTTACGCTCATTAACCTATGACCTGACAGAGGAGCGCTTGAAGAGTCCCAAGACG 5280  
 QY 2276 ProIleIleAsnThrProValValSerIleSerValHisAspAspGlyLysLeuPro 2295

DB 5281 CCGATCATCAACACCCGCTGGAGCATCAGGTCCATGATGATAGAGAGCTTCCGCC 5340  
 QY 2296 ArgAlaLeuAspLysProValThrValGlnPheArgLeuLeuGlnThrGlyLysArgThr 2315  
 DB 5341 CGGGCCCTGAGCAAAACCCCTCAAGGTGCTCCGCTCTGAGACAGAGAGCGGAGCC 5400  
 QY 2316 LysProIleCysValPheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSer 2335  
 DB 5401 AAGCCATGTGTCTTCTTGTGAACCACTCAATCTGTGTGATGAGCAAGGTGCTGTG 5460  
 QY 2336 AlaArgGlyCysGluValValPheArgAsnGlySerHisValSerCysGlnCysAsnHis 2355  
 DB 5461 GCCAGAGGCTGAGAGTGGCTTCCGACATGAGAGCACTCAGGCTGCCAGTCAACAC 5520  
 QY 2356 MetThrSerPheAlaValLeuMetAspValSerArgArgGlyLysGlnIleLeuPro 2375  
 DB 5521 ATGACGAGCTTGGCTGTGTCTATGACGTTTCTCGCGGAGAAATGGGAGATCTTGCA 5580  
 QY 2376 LeuLysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhe 2395  
 DB 5581 CTGAGACACTGACATACGTGGCTCTAGGTGTCACTTGGCTGCCCTTGTGCTCACTTC 5640  
 QY 2396 PhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeu 2415  
 DB 5641 TCTTCTCACTCTTGGGTATCTTGCGCTCAACCAACAGGATCCGAGCTAAGCTG 5700  
 QY 2416 ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeu 2435  
 DB 5701 ACAGCTGCTCTGGCTGGCTCAGCTGAGTCTTCTCTGGGAATCAACAGGCTGACCTTC 5760  
 QY 2436 PropheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSer 2455  
 DB 5761 CTTTTGCTGACAGTATTCATTCAGCTCTGCTGCACTTCTGATCCTTGACACTTTTCC 5820  
 QY 2456 TyrAlaLeuLeuGlnAlaLeuHisLysTyrArgAlaLeuThrGlyLysArgAspValAsn 2475  
 DB 5821 TGGGCTCTGAGAGGCTTGCACCTGTACCGGAGCACTCAGAGGTGGCGCATGTCAAC 5880  
 QY 2476 ThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGly 2495  
 DB 5881 ACCGGCCCATCGCTTCACTCAACATGTGGGCTGGGGGTGTGCTGCTTCAACAGGG 5940  
 QY 2496 LeuAlaValAlaGlyLeuAspProGlyLysTyrGlyAsnProAspPheCysTyrLeuSerIle 2515  
 DB 5941 CTAGCGGTGGGCTGGACCCGAGGGCTACGGGAACCCGACTTTCGTGGCTTCCATC 6000  
 QY 2516 TyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSerVal 2535  
 DB 6001 TATGACACGCTCATCTGAGAGTGTGCTGGCCGCGTGGCCGTTCGATGAGTGTTC 6060  
 QY 2536 PheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyLys 2555  
 DB 6061 TTCTGTATCATCTGGGCGGCGCGGCTCTGTGTGCTCCACAGGGAGGAGCTTTGAGAGG 6120  
 QY 2556 LysGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThr 2575  
 DB 6121 AAAGGTCCGTCTCGGGCTGAGGCTCTCTTCCGCGTCTCTGTGTGAGCGGCAAG 6180  
 QY 2576 TyrLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAla 2595  
 DB 6181 TGGCTGTGGCACTGCTCTGTCAACAGCGACACCTCTCTTCCATCACTCTTGTGCT 6240  
 QY 2596 ThrCysAsnCysAlaGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlu 2615  
 DB 6241 ACTGCAATTTGCATCCAGGGCCCTTCACTTCTCTCTATGTGTGCTTGAAGAGG 6300  
 QY 2616 ValArgLysAlaLeuLysLeuAlaCysSerArgLysProSerProAspProAlaLeuThr 2635  
 DB 6301 GTCCGGAAAGCACTCAGCTTCTGTGACGCGAGCCAGGCCGCTGACCTCTGACAC 6360  
 QY 2636 ThrLysSerThrLeuThrSerSerTyrAsnCysProSerProTyrAlaAspGlyLysLeu 2655

Db	6361	ACCAAGTCACCCGACCTCGTCCCTACAGACTGCCCCAGCCCTTACGCAAGATGGGCGGCTG	6420
Qy	2656	TyTgInP-roIyTgIyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSer	2675
Db	6421	TACACGACCTTACGGAGACTCGGCGGCTCTCTGCAACAGCACAGTCGCTCGGCGAAGAT	6480
Qy	2676	GlnProSerTyrIleProPheLeuLeuArgGlnGlnSerAlaLeuAspProGlyGlnGly	2695
Db	6481	CAGCCCAAGCTACATCCCTTCTTGCTGAGGAGAGATCCGACTGAAACCTTGGCCAAAGG	6540
Qy	2696	ProProGlyLeuGlyAspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAsp	2715
Db	6541	CCCCCTGGCCCTGGGGGATCCAGAGCAGCTGTCTTCCTGAAAGGTCAAGACACAGACATGAT	6600
Qy	2716	ProAspThrAspSerAspSerAspLeuSerLeuGlnAspArgGlnSerGlySerTyrAla	2735
Db	6601	CCTGACACGGACTCCGACAGTGAAGCTGTCTCTTAAAGACGACACAGAGTGGTCTTAAAGC	6660
Qy	2736	SerThrHisSerSerAspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaPhe	2755
Db	6661	TCTAACCACTATCAGACAGTGAAGAGGAAAGAAAGAGAGAGAGAGAGAGCGGCTTC	6720
Qy	2756	ProGlyGlnGlnGlyTyrPhePheSerLeuLeuGlyProGlyAlaGlnIuArgLeuProLeuHis	2775
Db	6721	CCTGAGAGCGAGGCTGGGATAGCTGCTGCGGCGCTGACAGAGAGATCGCCCTTCAC	6780
Qy	2776	SerThrProLysAspGlnGlyProGlyLysAlaProTyrProGlyLysPheGly	2795
Db	6781	AGTACTCCCAAGATGGAGGGGCCCAAGGGCTCTGGCAAGGCCCTTGGCCAGAGACTTGGG	6840
Qy	2796	ThrThrAlaLysGlnSerSerGlyLysAsnGlyAlaProGlnGlnIuArgLeuArgGlnAsnGly	2815
Db	6841	ACCACAGCAAAAGAGAGTATGTGGCACAGGGGCCCTGAGGAGCGGCTGGGGAATAGGA	6900
Qy	2816	AspAlaLeuSerAspArgGlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHis	2835
Db	6901	GATGCCCTGTCTCAGAGGGGTCTCCTAGGGCCCCCTTCAAGGCTCTTCTGGCCACGCTCAC	6960
Qy	2836	LysGlyIleLeuLysLysLysCysLeuLeuProThrIleSerGlnLysSerSerLeuLeuArg	2855
Db	6961	AAAGGCAATCCTTAAAGAGAACTGTCTGCCACATCAGCAGCAAGAGACAGCTCTCTCGG	7020
Qy	2856	LeuProLeuGlnGlnGlyThrGlySerSerArgGlySerSerAlaSerGlnGlySerArg	2875
Db	7021	CTCCCCCTGGAGCAATGACAGGGTCTTCCCGGGGCTCTCCGCTAATGTAGGGCAGCCGG	7080
Qy	2876	GlyGlyProProProAlaArgProProProArgGlnSerLeuGlnGlnGlnLeuAsnGlyVal	2895
Db	7081	GGCGGGCCCCCTCCCCGCGCACCGCCCCGGGAGAGCTCCAGAGGACAGCTGAACGGGGTTC	7140
Qy	2896	MetProIleAlaMetSerIleLysAlaGlyThrValAspGlnAspSerSerGlySerGln	2915
Db	7141	ATGCCCAATCGGCATGAGATCAAGGACGAGCAGCGGTGATGAGGACTGTGACGGCTCCGAA	7200
Qy	2916	PheLeuPhePheAsnPheLeuHis	2923
Db	7201	TTTCTCTCTTAACTTCTGGAT	7224

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AX367102	AX367102	Sequence 21 from Patent WO0198354.	AX367102	AX367102	GI:18855304	Homo sapiens (human)	Homo sapiens		
						Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
								1	Griffin, J.A., Kallick, D.A., Tribouley, C.M., Yue, H., Nguyen, D.B., Tang, T.T., Lai, P., Policky, J.L., Azimzal, Y., Lu, D.A., Gaul, R., Yao, M.G., Buford, N., Hafalia, A.J., Baughn, M.R., Baidman, O.

```

TITLE
JOURNAL
Paterson, C., Yang, J., Xu, Y., Warren, B.A., Ding, L. and
Sanjanmala, M.S.
Receptors
Patent: WO 0198354-A 21 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..9321
/organism="Homo sapiens"
/mol_type="genomic DNA"
/dbi_xref="taxon:9606"
/note="Incyte ID No: 6977010CB1"
BASE COUNT
1780 a 3008 c 2776 g 1757 t
ORIGIN

```

Alignment Scores:	
Pred. No.:	0
Score:	1958.00
Percent Similarity:	96.91%
Best Local Similarity:	96.91%
Query Match:	66.99%
DB:	6
Length:	9322
Matches:	2886
Conservative:	0
Mismatches:	17
Indels:	92
Gaps:	0

US-09-916-849A-3 (1-2923) x AX367102 (1-9321)

[illegible]

Db 982 AACACCCAGTCTTCAAGGCTCAGCGCGCAGAGCAAGCATGCCCCGAGAAAGTCCCTG 1041  
 Qy 273 AATThleuThrlleuValThraapThraaaphiAaapProValPhegluIn 292  
 Db 1042 GCTACCTCAACCACTTGTGTTACTGACCAATGACCATGACCTTGTGTCAGCGACG 1101  
 Qy 293 GluTyZLyGluSerLeuArgGluAenLeuGluValGlyTyrgluValleuThraValArg 312  
 Db 1102 GAGTACAGAGAGAGCTCAGAGAGAACTGAGAGTTGGCTATAGAGTGTCTCATGTGCAAG 1161  
 Qy 313 AATThraapGlyAaapAaaproAaenAaenileuTyraArgLeuLeuGluGlySer 332  
 Db 1162 GCACAGAGATGATGATCCCTCCCAATGCAATATCTGTACCGCTGTGAGAGGAGTCT 1221  
 Qy 333 GlyGlySerProSerGluValPhegluValAaapProArgSerGlyValleuArg 352  
 Db 1222 GGGGGAGGCCCCCTTAAGTCTTGAAGTCGACCTGCTGAGGGTGAATCCGACCCGT 1281  
 Qy 353 GlyProValAaapArgGluGluValGluSerTyrgluLeuThraValGluAaapArgIn 372  
 Db 1282 GGGCCGTGGATCGGGAAGAGTGGAACTCTTACAGCTGACGGTACAGGCAATGACAG 1341  
 Qy 373 GlyArgAaapProGlyProArgSerThraValAaValPheLeuSerValGluAaap 392  
 Db 1342 GGTCCGAGACCCGGGTCTCGAGTACACAGCCGCTGTTTCTTCTGTGAGAGATGAC 1401  
 Qy 393 AaapAaapAaapAaaproInPheSerGluValArgTyraValGluValArgValAaapVal 412  
 Db 1402 AATGAATAAGTCCCCCAAGTTTAACTGAGAAAGCTAATGTTCCAGGTGAGAGGAGAGTGG 1461  
 Qy 413 ThrProGlyAaapProValleuArgValThraAaAaapArgAaapGlySerAaenAa 432  
 Db 1462 ACTCCAGAGGGGCCCAAGTACCGAGTCAACAGCTCCGATCGAACAAGGAGACATGCC 1521  
 Qy 433 ValValHisTySerIleMetSerGlyAaenAaArgGlyGluPheTyrieuAaapAaGln 452  
 Db 1522 GTGGTCACTATAGCATCATGAGTGGCAATGCTGGGGAGCATGTTTATCTGAGTGGCCAG 1581  
 Qy 453 ThrGlyAaLeuAaapValValSerProLeuAaapTyrgluThraValArgGluTyraLeu 472  
 Db 1582 ACTGGAGCTTGAATGTGTGAGCCCTCTTGACTATGAGACGACCAAGAGTACACCTTA 1641  
 Qy 473 ArgValArgAaGlnAaapGlyArgProProLeuSerAaenAaAaSerGlyLeuValThra 492  
 Db 1642 CGGGTGGAGCAAGAGTGTGCGCCGCCCCCACTCTTAATGTCTGTGCTTGGTGGAGCA 1701  
 Qy 493 ValGlnValleuAaapIleAaenAaAaenAaAaProIlePheValSerThraProPheGlnAa 512  
 Db 1702 GTACAGGTCTGTGATATCAACGACAAATGCCCATCTTCGTACAGCAACCCCTTCCAGGCT 1761  
 Qy 513 ThrValleuGluSerValProLeuGlyTyraLeuValleuHisValGlnAaIleAaapAa 532  
 Db 1762 ACTGTCTGAGAGAGCTCCCTTAAGCTTACCTGTGTTCCATGTCCAGGCTTACAGAGCT 1821  
 Qy 533 AaapAaGlyAaAaenAaAaArgLeuGluTyraArgLeuAaGlyValGlyAaAaPhePro 552  
 Db 1822 GATGCTGTGACAAATGCGCCCTGGAATACCGCTTCTGAGGGGAGCAATGACTTCCCC 1881  
 Qy 553 PheThraIleAaenAaGlyThraGlyTyraIleSerValAaAaGluLeuAaAaArgGluGlu 572  
 Db 1882 TTCAACATCAACAATGACAGGCTGATCTGTGTGCTGTGAATCTGACCGAGAGAA 1941  
 Qy 573 ValAaPheTyraSerPheGlyValGluAaAaArgAaAaGlyThraProAaLeuThraAa 592  
 Db 1942 GTTGATTTCTACAGCTTTGGGGTGAAGCTCGAGACCAATGCACTCAAGACTCACTGCC 2001  
 Qy 593 SerAaAaSerValSerValThraValleuAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 612  
 Db 2002 TCGGCGCATGTACAGCTGTCTGTGATGTCAACCAATCAATCAATCAATCAATCAATCA 2061  
 Qy 613 ProGluTyraThraValAaGluAaenGluAaapAaAaAaAaAaAaAaAaAaAaAaAaAaAa 632

Db 2062 CCAGAGTACACAGTGGCTCATAGATGAGATGCACTGTGGGACCAAGGTGTGAGCGGTG 2121  
 Qy 633 SerAa 652  
 Db 2122 TCAAGCTGAGACCGGATGCTCATAGTGTCAATCACTTACAGATCAACAGTGGCAATACT 2181  
 Qy 653 ArgAaenAaPheSerIleThraSerGlnSerGlyGlyGlyLeuValSerLeuAaLeuPro 672  
 Db 2182 CGAAACCCCTTCTCATCAACAGCAAGCAAGTGTGTGGCTGTGTATCCCTTCCCTGCCA 2241  
 Qy 673 LeuAaPtyrGlyLeuGluArgGlnTyraValleuAaValThraAaSerAaapGlyThraArg 692  
 Db 2242 CTGAGCTCAAACTTACAGCGGAGATGATGTGGTGTGTATCCCTCCGATGAGCACTCGG 2301  
 Qy 693 GlnAaPThraAaGlnIleValAaAaValThraAaAaenThraAaAaProValPhe 712  
 Db 2302 CAGAGACGGCACAAATGTGTGATGTCAACGAGCGCAACCAACCAATGCTTCTTCTT 2361  
 Qy 713 GlnSerSerHisTyraThraValAaenValAaenAaAaAaAaAaAaAaAaAaAaAaAaAa 732  
 Db 2362 CAGAGCTCCCACTATACAGTGAATGTATAGAGACCGGCGGAGAGCAACGAGTGGT 2421  
 Qy 733 LeuIleSerAaAaThraAaPgluAaPThraGlyGluAaenAaAaGlyIleThraTyraPheMetGlu 752  
 Db 2422 CTGATCAAGGCCACGAGTGAAGACACAGTGAATGCCCCGATCACTTCAATGAG 2481  
 Qy 753 AaPserIleProGlnPheArgIleAaapAaAaPThraGlyAaValThraGlnAaGlu 772  
 Db 2482 GACAGATCCCACTTCCGATTCGATCAATGACAGAGGGGCTGTACACCAAGCTGAG 2541  
 Qy 773 LeuAaPtyrGluAaPgluAa 792  
 Db 2542 CTGAGCTACAGAGACCAAGTGTCTTACACCTGCGCATTAAGTCTCGGAGCAATGCAAT 2601  
 Qy 793 ProGlnTyraSerAaPThraThraTyraLeuGluIleLeuValAaAaAaAaAaAaAaAaAa 812  
 Db 2602 CCCAGAAAGTCCGACACCACTTACCTGAGATCCGTGGAGAAAGAGTGAATACAAATGCC 2661  
 Qy 813 ProGlnPheLeuAaAaPserTyrgluGlySerValTyrgluAaAaAaAaAaAaAaAaAa 832  
 Db 2662 CTTAGTCTTCCGAGACCTCCACAGGGCAAGTGTATGAGAGTGTGCACTTCACT 2721  
 Qy 833 SerValleuGlnIleSerAaAaThraAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 852  
 Db 2722 ACGTCTCGCAATCTCAAGCACTGATCGTGAATCTGAGCTTAATAGCGAGGCTCTTAC 2781  
 Qy 853 ThrPheGlnGlyAa 872  
 Db 2782 ACCCTTCAAGAGGCGAGCATGAGAGCGTGAATTTATGTGTGAGTCAAGTCAAGGCAATC 2841  
 Qy 873 ValArgThraLeuArgArgLeuAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 892  
 Db 2842 GTGGCAAGCTTACAGAGCTGTGATCGAGAGACGTGGCCCAATATGTCTTGGGGCATAT 2901  
 Qy 893 Aa 912  
 Db 2902 GCAAGTGAACAGAGGATGCCCCCAGCGGCAACCTATGAGAGTGAATCACTGTGTG 2961  
 Qy 913 AaPValAa 932  
 Db 2962 GATGTGAATGACATCCCTCTGTGTGAGCAAGATGATGATGTGTGTGTGTGAGAGAG 3021  
 Qy 933 AaenSerProIleGlyLeuAa 952  
 Db 3022 AACAGCCCATTTGGGTGACCGTGGCGGGTCAAGCAGCACTGACCCGATGAGAGGCAAC 3081  
 Qy 953 AaenAaGlnIleMetTyraGlnIleValGluGluAaenIleProGluValPheGlnLeuAaAa 972  
 Db 3082 AATGCCCAATTAATGACAAATGTGTGAGGAGCAATCCCTGAGAGTCTTCCAGCTGAGAC 3141  
 Qy 973 IlePheSerGlyGluLeuThraAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 992  
 Db 3142 ATCTTCTCCGGGAGCTGACAGCCCTGTGATGATGATGATGATGATGATGATGATGATGAT 3201

QY 993 ValLeuVal11eglnAlaThrSerAlaProLeuValSerArgAlaThrValHISValArg 1012  
DB 3202 GTCTGTCATCCAGGCCAGCTCAAGCTCTCTGCTGAGCCGGGCTACAGTCCAGCTCCG 3261  
QY 1013 LeuLeuAspArgAsnAspAsnProProValLeuGluYAsnPhgGlu11LeuPheAsnAn 1032  
DB 3262 CTCTCTGACCGCAATGACAAACCAACGAGCTGGGCACTTGAGATCTCTTTCAACAAC 3321  
QY 1033 TyrVal1ThrAsnArgSerSerPheProGluYAla11eg1YArgValProAlaHIS 1052  
DB 3322 TATGTCACCAATCGCTCAAGAGCTTCCCTGGGGGCGCATTTGGCGGATCTGCGCAT 3381  
QY 1053 AspProAsp11SerAspSerLeuThrTyrSerPheGluArgGluYAsnGluLeuSerLeu 1072  
DB 3382 GACCTGAAATCTCAAGATGCTGACTTACAGCTTGAACGGGGAATACCTCAAGCTCG 3441  
QY 1073 ValLeuLeuAsnAlaSerThrGluLeuYLeuSerArgAlaLeuAspAsnAnArg 1092  
DB 3442 GTCTGCTCAATGCTTCAACGGGTGAGCTGAAGCTGAAGCGCGCACTGGAACAACCGG 3501  
QY 1093 ProLeuGluAla11eMetSerValLeuValSerAspGluYAlaHISerVal1ThrAlaGln 1112  
DB 3502 CCTCGAGGCGCATATAGCGGTGCTGCTGAGACGGGTGACAGCGCTGACCGCCAG 3561  
QY 1113 CysAlaLeuArgVal1Thr11e11eThrAspGluMetLeuThrHISer11eThrLeuArg 1132  
DB 3562 TGGCGGCTGCTGCTGATCATCAACGATGAGATGCTCAACCAAGCATCAAGCTGCGC 3621  
QY 1133 LeuGluAspMetSerProGluArgPheLeuSerProLeuLeuGluYLeuPhe11eglnAla 1152  
DB 3622 CTGAGAGACATGTCACCGAGGGCTTCCTGTCACCACTGAGCGCTTCTCATCCAGGG 3681  
QY 1153 ValAlaAla1ThrLeuAla1ThrProProAspHISValVal1PheAsnValGlnArg 1172  
DB 3682 GTGGCGGCAAGCTGGCCACCGCAACCGGACCAAGTGTGTCTTCAACGTACAGCGGAG 3741  
QY 1173 ThrAspAlaProGluYAla11e11eLeuAsnValSerLeuSerValGluYLeuProGluY 1192  
DB 3742 ACCGAGCGCCCGGGGGGCACTCTCAACGTGAGCTGTCTGCTGGGCGACGCGCAGGG 3801  
QY 1193 ProGluYAlaYProProPheLeuProSerGluAspLeuGlnArgLeuThrLeuAn 1212  
DB 3802 CCGGGGGGGGGCGCGCTTCTCTGCTGAGACCTGCAAGAGGCGCTATACCTCAAC 3861  
QY 1213 ArgSerLeuLeuThrAla11eSerAlaGlnArgValLeuProPheAspAsn11eCys 1232  
DB 3862 CGCAAGCTGCTGACGGCCATCTCGGCAACGGCGGTGTGCTTGAAGCAACAATTCTGC 3921  
QY 1233 LeuArgGluProCysGluAsnTyrMetArgCysValSerValLeuArgPheAspSer 1252  
DB 3922 CTGCGGAGGCGCTGCGAGAACTACATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3981  
QY 1253 AlaProPhe11eAlaSerSerSerValLeuPheArgPro11eHISProValGluYLeu 1272  
DB 3982 GCGCCCTTCATGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4041  
QY 1273 ArgCysArgCysProProGluYPheThrGluYAspTyrCysGluThrGluValAspLeuCys 1292  
DB 4042 CGCTCGCGCTGCGCGCGGCTTCAAGGCTGACTCTGAGAACCGAGGTGAGACTTCTGC 4101  
QY 1293 TyrSerArgProCysGluYProHISGluYArgCysArgSerArgGluYAlaYThrCys 1312  
DB 4102 TACTGCGCGGCTCTGAGCGCCCAACGGGCGCTGCGCGAGCGCGAGGGGCGCTACACTGC 4161  
QY 1313 LeuCysArgAspGluYThrGluYHISGysGluValSerAlaArgSerGluYArgCys 1332  
DB 4162 CTCTGCTGATGAGCTACAGGAGTGAACACTGAGAGTGAAGTGTCTGCTGAGCGGTGCG 4221  
QY 1333 ThrProGluYAlaCysLeuAsnGluYAlaYThrCysValAsnLeuLeuValGluYLeuPheCys 1352  
DB 4222 ACCCGGGGTGTCTGCAAGAAATGGGGGCACTGTGTCAACTGCTGCTGCTGCGGCTTCAAG 4281

QY 1353 CysAspCysProSerGluYAspPheGluYsProTyrCysGlnVal1ThrThrArgSerPhe 1372  
DB 4282 TGCATGTCCTCATCTGAGAACTTGCAGAAAGCCCTTCTGCGAGGTGACCAACGGGAGCTTC 4341  
QY 1373 ProAlaHISerPhe11eThrPheArgGluYLeuArgGlnArgPheHISerThrLeuAla 1392  
DB 4342 CCGGCGCACTCTTCATACCTTTTGGCGGCTGCGCGCGCTTTCACCTTCACTTCCGCGC 4401  
QY 1393 LeuSerPheAla1ThrLeuGluYAspGluYLeuLeuLeuYArgGluYArgPheAsnGlu 1412  
DB 4402 CTCTGCTTTCACAAAGAGGCGCAACGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4461  
QY 1413 LysHISAspPheValAlaLeuGluVal11eglnGluValGlnLeuThrPheSerAla 1432  
DB 4462 AAGCATGACTTGTGTGGCCCTCGAGGTGATCCAGAGAGCTTCAAGCTCACTTCTTGA 4521  
QY 1433 GluYLeuSerThrThrThrValSerProPheValProGluYAlaSerAspGluYLeuTTP 1452  
DB 4522 GGGAGTCAACCAACGAGGTGCTCCCATTCGTGCGCGAGAGGTGATGATGGCCAGTGG 4581  
QY 1453 HISThrValGlnLeuYsTyrTyrAsnLysProLeuLeuGluYGlnThrGluYLeuProGln 1472  
DB 4582 CATACGTGCGAGCTGAATATCTACAAATAAGCACTTTGGGTGAGACAGGCTCCCAAG 4641  
QY 1473 GlYProSerGluGlnYsValAlaVal1ThrValAspGluYCysAspThrGluYAla 1492  
DB 4642 GGCCCATCAGAGAGAGAGGTGGTGTGTGACCGGTGATGCTGTGACACAGAGATGGCC 4701  
QY 1493 LeuArgPheGluYsValLeuGluYAsnTyrSerCysAlaAlaGlnGluYThrGlnGluY 1512  
DB 4702 TTGGCTTCGAGTCTGTCTGGCACTATCTCTGTGTGCTGCGCGGCAACCAAGGTGAG 4761  
QY 1513 SerYsLysSerLeuAspLeuThrGluYProLeuLeuLeuGluYAlaProAspLeuPro 1532  
DB 4762 AGCAAGAGTCTGTGANTCTGACCGGGGCCCTGTCTACAGCGGGGTCTGACTGCTGC 4821  
QY 1533 GluSerPheProValArgMetArgGlnPheValGluYCysMetArgAsnLeuGlnValAsp 1552  
DB 4822 GAGAGCTTCCAGTCCGATCGGATCGGAGTCTGTGGCTGCAAGCGGAACTGCAAGGTGAC 4881  
QY 1553 SerArgHIS11eAspMetAlaAspPhe11eAlaAsnAnGluYThrValProGluYsPro 1572  
DB 4882 AGCGGCAATAGACATGAGCTGATCTTCAATGCAACATGAGCACTGCTGCTGCTGCT 4941  
QY 1573 AlaYsLysAsnValCysAspSerAsnThrCysHISAsnGluYAlaYThrCysValAsnGln 1592  
DB 4942 GCAAGAGAGAGGTGTGAGACAGCAACTTGCACAAATGGGGGCACTTGGTGAACAG 5001  
QY 1593 TTPAspAlaPheSerCysGluYsProLeuGluYPheGluYAlaYsSerCysAlaGlnGlu 1612  
DB 5002 TGGAGCGGTTCACTGCAAGTGCCTCCCTGGGCTTGGGGGCAAGAGCTGCGCCAGAA 5061  
QY 1613 MetAlaAsnProGluYHISPheLeuGluYsSerSerLeuValAlaTPHISGluYLeuSerLeu 1632  
DB 5062 ATGGCAATCAACAGCACTTCTGAGAGAGCGCTGAGTGGCTGAGCAAGGCTCTGCTG 5121  
QY 1633 Pro11eSerGlnProTyrTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGluYAla 1652  
DB 5122 CCAATCTCCCAACCTGTGACTCAAGCTCATGTTCCGCAAGCGCCAGCGGAGGTGTC 5181  
QY 1653 LeuLeuGlnAla11eThrArgGluYArgSerThr11eThrLeuGlnLeuArgGluYHIS 1672  
DB 5182 CTGCTGAGGCGCAACCAAGGGGCGAGCACTCAACCTTCAAGCTTCAAGAGGGCGAC 5241  
QY 1673 ValMetLeuSerValGluYAlaYThrGluYLeuGlnAlaSerSerLeuArgLeuGluYProGluY 1692  
DB 5242 GTGATGCTGAGCGTGAAGGGCAACAGGCTTCAAGGCTTCTCTCTGCTGAGCGCAAGC 5301  
QY 1693 ArgAlaAsnAspGluYAspTTPHIS11eAlaGlnLeuAlaLeuGluYAlaSerGluYPro 1712  
DB 5302 CGGCGCAATAGCGGTGACTGGCAACATGCAACAGCTGGGCACTGGAGCGAGGGGGGCT 5361  
QY 1713 GlyHISAla11eLeuSerPheAspTyrGluYLeuGlnArgAlaGluYAsnLeuGluYPro 1732

Db 5362 GGGCAAGCCATTCCTGCTCCGATTATGGGCAAGAGAGAGAGGCAACCTGGGCCCC 5421  
 Qy ArgLeuHieglYleuHiebleuSerAmiIleThrValGlglyIleProglYProAlaGly 1752  
 Db 5422 GGGCTCAAGGTGTGGACCTGAGCAACATACAGTGGCGGAATACCTGGGCGCAAGCGGG 5481  
 Qy 1753 GilyValAlaArgGlyPheArgGlyCybleuGlnGlyValArgValSerAspThrProGlu 1772  
 Db 5482 GGTGTGGCCCGTGGCTTTGGGGCTGTGGTTCAGGGTGTGGGGTGGAGCGATACCGCAAG 5541  
 Qy 1773 GilyValAlaSerIleuAspProSerHieglYgluSerIleAsnValGluGlnGlyCySer 1792  
 Db 5542 GGGGTTCACAGCTGATCCACAGCATGGGAGAGATCAACCTGAGAGAGAGCTGTACG 5601  
 Qy 1793 LeuProAspProCyAspSerAspProCyProAlaAsnSerTyCySerAspAspThr 1812  
 Db 5602 CTGCTGACCTTGTGATCAACCGGTCTGTGCTACAGCTATGTGACAGCAAGACTGG 5661  
 Qy 1813 AspSerTySerCySerCyAspProGlyTyTyTyArgValAspAsnCySerAspVal 1832  
 Db 5662 GACAGCTATTCTGACAGCTGTATCCAGGTTCATAGGTGACAACTGTACTATGTGTGT 5721  
 Qy 1833 AspLeuAspProCySerGlyHieglYgluSerValCyThrArgIleProSerAlaProHieglY 1852  
 Db 5722 GACCTGAACCCGTGTAGCAACAGCTGTGTATCCGCAAGCCAGTGGCCCCCATGGC 5781  
 Qy 1853 TyThrCySerGluCyAspProAspTyIleuGlyProTyCySerGluThrArgIleAspGln 1872  
 Db 5782 TATACCTGCAAGTGTCCCGCAAAATACCTTGGGCAATCTGTAGACCAAGATTTAGCAAG 5841  
 Qy 1873 ProCyAspProArgIlyTyTyTyArgIleAspProThrCySerGlyProCyAspAspValSer 1892  
 Db 5842 CTTGTGCCCCGTGGCTGTGGTGGGACATCCCAATGTGGCCCATGCAACTGTATGTGTCAAG 5901  
 Qy 1893 IysGlyPheAspProAspProAspValGlyHieglYgluSerGlyGlyGlyGlyGlyGly 1912  
 Db 5902 AAAGGCTTTGACCAACATGCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5961  
 Qy 1913 TyArgProProGlySerProThrCybleuLeuCyAspProAspTyProThrArgIleSerIleu 1932  
 Db 5962 TACCGGCCCCAG 6021  
 Qy 1933 SerArgValCyAspProGluAspGlyGlnCyAspProCyAspProGlyValIleGlyArg 1952  
 Db 6022 TCCAGAGTCTGTACCTGAGATGGCAGGTGTCCATGCAAGCCAGAGTGCATCGGGCGT 6081  
 Qy 1953 GlnCyAspArgCyAspAspProPheAlaGluValThrThrArgGlyCySerGlu----- 1970  
 Db 6082 CAGTGTGACCGCTGTGACAACTTTGTGAGGTCAACCAATGAGCTGTGAGAGGGGCC 6141  
 Qy 1971 --ValAsnTyAspSer--CyAspProAlaIle-Glu-AlaGlyIleTyTyTyProArg 1988  
 Db 6142 TTTGTTTG-CT-AG-TT-ACGTGTCCCG-AGCCA-TAGAGTGTGGCC--TC-CA-AG-AG 6190  
 Qy 1988 gthr-----Arg-----PheGly-----LeuProAla-AlaAlaProCyAsp 2000  
 Db 6191 AACCTGACAGAGCTGACAGGGGCTTCCT-GTGTGTCTCCCTGA-GAGCGG-CCCTT-TT 6246  
 Qy 2000 olysglySerPhe-----GlyThrAlaValArgHieCyAspGlnHieArgGlyTyTyTyPleu 2018  
 Db 6247 -GG-CT-TCTTTCCTCCCAAGGAGCTGTGTGGCTGACCTGTATAGAGCAAGGGGTGAGCTC 6303  
 Qy 2019 ProProAsnLeuPheAsnCySerIleThrPheSerGluLeuTyAspGlyPheAlaGlu 2038  
 Db 6304 CCCCACAACTCTTCAACTGCACTGCATCACTTCTCAAGATGAGAGGCTTCGTGAG 6363  
 Qy 2039 ArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeu 2058  
 Db 6364 CGGCTACAGGGAGATGTGAGGCTAGACTCAGGGGCTCCACAGAGAGTACGCTGCTC 6423  
 Qy 2059 LeuArgAsnAlaThrGlnHieThrAlaGlyTyTyPheGlySerAspValIysValAlaTy 2078

Db 6424 CTGGCAACGCCACGCAAGCAACAGCTGTACTTGGCAGGCAAGTCAAGTGGCTTAC 6483  
 Qy 2079 GlnLeuAlaThrArgLeuLeuAlaHieglYgluSerThrGlnArgGlyPheGlyLeuSerAla 2098  
 Db 6484 CAGTGGCCACCGGCTGTGTGGCCCAAGAGACCCAGGGGGCTTTGGGCTGTGTGCTCC 6543  
 Qy 2099 ThrGlnAspValHiePheThrGluAsnLeuAspArgValGlySerAlaLeuLeuAspThr 2118  
 Db 6544 ACACAGAGAGTCACTTCACTGAGAAATGTGCTGGCGGGTGGGAGGCGCTCTGTGACACA 6603  
 Qy 2119 AlaAsnValArgHieTyTyTyGluLeuIleGlnGlnThrGlnGlyGlyTyThrAlaThrPleu 2138  
 Db 6604 GCCAACAGCGGCACTGGAGCTGATCCAGACAGAGAGGGTGGACCGCTGTGTGCTC 6663  
 Qy 2139 GlnHieTyTyTyAlaTyAlaSerAlaLeuAlaGlnAspMetArgHieThrTyTyTySer 2158  
 Db 6664 CAGCACTTAGAGCTTACCGCAAGTGCCTTGGCCCAAGATCCGCGCAACCTTACCTTACG 6723  
 Qy 2159 ProPheThrIleValThrProAsnIleValIleSerValAlaArgLeuAspIysGlyAsn 2178  
 Db 6724 CCTTCAACATGTGACCGCCCAACATGTTCATCTCCGTAGTGGCTTGGACAAAGGAAAC 6783  
 Qy 2179 PheAlaGlyAlaIysLeuProArgTyTyTyGluAlaLeuArgGlyGluGlnProProAspLeu 2198  
 Db 6784 TTTGTGGGGCCAACTGCCCCGCTTACAGAGCGCTGTGGAGAGAGAGAGAGAGAGAGAG 6843  
 Qy 2199 GluThrThrValIleLeuProGluSerValPheArgGluThrProProValAlaArgPro 2218  
 Db 6844 GAGACAAAGTCAATCTGTGCTGATGTCTGTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 6903  
 Qy 2219 AlaGlyProGlyGlyAlaGlnGluProGluLeuAlaArgArgIleArgArgHiePro 2238  
 Db 6904 GAGAGCCCCGAGAGAGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6963  
 Qy 2239 GluLeuSerGlnGlyAlaValAlaIleSerValIleIleTyArgThrLeuAlaGlyLeu 2258  
 Db 6964 GAGTGTAGCCAGAGGTGAGGCTGTGGCAGGTCATCATCTACCTGAGCCGAGGCTA 7023  
 Qy 2259 LeuProHieAsnTyAspProAspIysArgSerIleAspValProGlyArgProIleIle 2278  
 Db 7024 CTGCTCACTTACTATGACCTTGTGACAGGAGCTTGAGAGTCCCAAGCCGCAATCATTC 7083  
 Qy 2279 AsnThrProValValSerIleSerValHieAspAspGluGluLeuLeuProArgAlaLeu 2298  
 Db 7084 AACACACCCGTGTGAGATCAGCTGCATGATGATGAGAGCTTCTGCCCCGGGCGCTG 7143  
 Qy 2299 AspIysProValThrValGlnPheArgLeuLeuGluThrGlnGluArgThrIlyAspIle 2318  
 Db 7144 GACAAACCGGTCAAGGTGAGTTCGGCTGTGAGACAGAGAGAGAGAGAGAGAGAGAGAG 7203  
 Qy 2319 CyValPheThrAsnHieSerIleLeuValSerGlyTyThrGlyGlyTyTyTySerAlaArgGly 2338  
 Db 7204 TGTGTCTTGTGAACTTCAATCTGTGATCTGTGATGAGAGAGAGAGAGAGAGAGAGAGAG 7263  
 Qy 2339 CySerGluValValPheArgAsnGlySerHieValSerCySerGlnCyAsnHieMetThrSer 2358  
 Db 7264 TGTGAAAGTCTTCCGCAATGAGAGCAAGCTGCAAGTGCAGAGTGCAGATGACAGAGC 7323  
 Qy 2359 PheAlaValLeuMetAspValSerArgArgIle----- 2369  
 Db 7324 TTTGCTGTGCTCATGAGAGTTCCTGGCGGAGAGTCCGGGAGAGAGAGAGAGAGAGAGAG 7383  
 Qy 2370 -----AsnGlyGluIleLeuProLeuTyThrLeuThrTyValAlaLeuGlyValAlaThr 2387  
 Db 7384 CCGTGAATGGGAGATCTGTGCACTGAAGACACTGATAGTGTGCTGAGTGTCAAC 7443  
 Qy 2388 LeuAlaAlaLeuLeuLeuThrPhePheLeuThrIleLeuLeuArgIleAlaLeuArgSerLeu 2407  
 Db 7444 TTTGCTGCTTGTGTCTCACTTCTTCTCTCACTTCTTGTGCTGTGCTGTGCTGTCAAC 7503  
 Qy 2408 GlnHieglYIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu 2427  
 Db 7504 CAACACGGATCCAGATCACTGACAGTCCCTGGGCTGTGCTCACTGCTCTTCTC 7563



QY 2428 LeuGlyTlaaenGlnAlaaspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis 2447  
 DB 7564 CTGGGAATCAACAGAGCTGACCTCCCTTGGCTGCAGAGTATTCGATCCGCTGCAC 7623  
 QY 2448 PheLeuTyrLeuCyThrPheSerTrpAlaLeuLeuGlnAlaLeuHisGLeuTyrAlaAla 2467  
 DB 7624 TTCCTGTACCTCTGCACCTTTTCTGAGCTGTGAGAGCTTGCAGCTGTAACCCGGGA 7683  
 QY 2468 LeuThrGlnValAlaArgAspValAsnThrGlyProMetArgPheTyrTrpMetLeuGlyTrp 2487  
 DB 7684 CTCACGAGAGTGGCGCATGTCAACACCGGCCCATGCTTCTACTACATGCTGGGCTGG 7743  
 QY 2488 GlnValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGlnGlyTyrGlyAsn 2507  
 DB 7744 GCGGTGCTGCTTCATCAACAGGCTAGCCGTGGGCTGGACCCCGAGGCTTACGGGAAC 7803  
 QY 2508 ProAspPheCysThrLeuSerIleTyrAspThrLeuIleTrpSerPheAlaGlyProVal 2527  
 DB 7804 CCTGACTTGTGCTGCTCTCCATCTATAGACGCTCATCGAGTTTCTGGCCCGGAG 7863  
 QY 2528 AlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla 2547  
 DB 7864 GCTTTGGCGGTCTGCATGAGTGTCTTCTGACATCTGGCGGCCCGGCTCTGTGCT 7923  
 QY 2548 AlaGlnArgGlnGlyPheGlnGlyGlyProValSerGlyLeuGlnProSerPheAla 2567  
 DB 7924 GCCCAGCGGCGAGGGCTTGTGAGAAAGGTCCTGTCGGGCTGCGAGCCCTTCCTGCGC 7983  
 QY 2568 ValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThr 2587  
 DB 7984 GTCTCTCTCTGTGAGCGCCAGTGGCTGTGGCACTCTCTGTCAACAGCGACAC 8043  
 QY 2588 LeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
 DB 8044 CTCCTCTTCACTACTCTTCTGTACTCTGCATGTGATTCAGGGCCCCCTTCACTTCTCTC 8103  
 QY 2608 SerTyrValValLeuSerIleGlyValAlaArgValAlaLeuLeuLeuAlaCysSerArgIle 2627  
 DB 8104 TCCCTATGTGTGTGTGAGCAAGGCTCCGAAAGCACTCAAGCTTCCGAGCGGCAAG 8163  
 QY 2628 ProSerProAspProAlaLeuThrThrIleThrIleThrLeuThrSerSerTyrAsnCysPro 2647  
 DB 8164 CCGAGCCCTGACCTGCTGTGACACAGTCCAGCTTACCTGCTGCTTCAACACTGCCCC 8223  
 QY 2648 SerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHis 2667  
 DB 8224 AGCCCTTACGAGATGGGGGCTGTACAGCCCTTACGAGACTCGGCCGCTCTCGCAC 8283  
 QY 2668 SerThrSerArgSerGlyIleSerGlnProSerTyrIleProPheLeuLeuArgGlnGln 2687  
 DB 8284 AACACCAAGTCCCTCGGGCAAGAGTCAAGCCAGCTACATCCCTTCTTCTGAGGAGGAG 8343  
 QY 2688 SerAlaLeuAsnProGlnGlnGlyProProGlnGlyLeuGlyAspProGlnGlySerLeuPheLeu 2707  
 DB 8344 TCCGCACTAAGCCCTGTGCAAGGCCCCCTGTGGCTGGGAGTCCAGGCAAGCTGTCTCTG 8403  
 QY 2708 GlnGlnGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGln 2727  
 DB 8404 GAAAGTCAAGACCAAGACATATCTGTACACGACTCCGACAGTACCTGTCTTGA 8463  
 QY 2728 AspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnGln 2747  
 DB 8464 GACGACCAAGTGGCTCTTATGCTTACCACTCATCGACAGCTAGAGAGAGAGAGAGAG 8523  
 QY 2748 GlnGlnGlnGlnGlnAlaAlaPheProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2767  
 DB 8524 GAG 8583  
 QY 2768 GlnAlaGlnArgLeuProLeuHisSerThrProTyrAspGlnGlnGlnGlnGlnGlnGln 2787  
 DB 8584 GAG 8643

QY 2788 AlaProTyrProGlnAspPheGlyThrThrAlaGlyGlnSerSerGlyAsnGlyValaPro 2807  
 DB 8644 GCCCCTGGCCAGAGAGACTTTGGAGCAACAGCAAAAGAGATGAGGCAACGGGCCCTT 8703  
 QY 2808 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2827  
 DB 8704 GAG 8763  
 QY 2828 ProGlnSerSerAlaGlnProHisGlyIleLeuLeuGlyLeuLeuGlyLeuProThrIle 2847  
 DB 8764 CCGAGCTTCTTGTCCACCTTCAACAGGCAATCTTATAGAGAGATGCTTCCCACTTC 8823  
 QY 2848 SerGlnGlnSerSerLeuLeuArgLeuProLeuGlnGlnGlnGlnGlnGlnGlnGln 2867  
 DB 8824 AGCGAG 8883  
 QY 2868 SerSerAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2887  
 DB 8884 TCTCTCGGTAGTGAAGGAG 8943  
 QY 2888 LeuGlnGlnGlnLeuAsnGlyValMetProIleAlaMetSerIleLeuAlaGlyThrVal 2907  
 DB 8944 CTCACGAG 9003  
 QY 2908 AspGlnAspSerSerGlySerGln 2915  
 DB 9004 GATGAG 9027  
 RESULT 7  
 QY 2908 AspGlnAspSerSerGlySerGln 2915  
 DB 9004 GATGAG 9027  
 LOCUS AK646739 24370 bp DNA linear PAT 04-MAR-2003  
 DEFINITION Sequence 931 from Patent EP1270724.  
 ACCESSION AK646739  
 VERSION AK646739.1 GI:28799122  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Suwa, M., Asai, K., Akiyama, Y. and Aburatani, H.  
 TITLE Guanosine triphosphate-binding protein coupled receptors  
 JOURNAL Patent: EP 1270724-A 931 02-JAN-2003;  
 National Institute of Advanced Industrial Science and Technology  
 (JIP) ; Center for Advanced Science and Technology Incubation, Ltd.  
 (JIP)  
 FEATURES  
 source Location/Qualifiers  
 1..24370  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 join1201..3510,8553..9200,11163..11385,11634..11749,  
 11929..12017,12408..12565,12927..13090,13263..13388,  
 13732..13898,14199..14486,14575..14749,14989..15130,  
 15312..15432,15856..16039,16226..16327,17669..17763,  
 17972..18181,18702..18885,19001..19129,19231..19411,  
 19545..19712,19815..19941,20054..20254,20547..20721,  
 21048..21160,21338..21417,21507..21633,21721..21843,  
 22399..22526,22761..22849,22954..23148,23287..23457,  
 23557..23791,24143..24170  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD69495.1"  
 /db\_xref="gi:28799123"  
 /translation="MRSAPATGVPLPTPEPPLILILILILPEPLIGDGVPCRSISGRS  
 RGSSGACAPMKLCPSSASNIMKLTYSKRDAGTILGILVPHNGARWCEBSAHNP  
 LPPAPBGPMSCRLIGTGHLSPOGKULPEHPCLEKAPRLSCSKAQAQGLRAGE  
 RSPRESIGGRKRVNNTAPOPSPSYATVBNPPAGTPVSLAIDPDEBARLEY  
 TMDALPDSRNOFSLDPVTGATVTAELBETSTVFRVTAADHGPRSLATLTL  
 ILVTDTHDHPVFOQKRYKGLRENLENGYVLTVRATDGPAPNANTLYLLSGSG  
 SPSEVPRI DPSSGVITRRGPDRREVSYOITVASAGDGRPRSTTAAPFLVEDD  
 NDNAPOGSEKRYVQVREDYTPGAPVYRATVTSDDKSNVYHVSINSNGARQGFYLD  
 AQTALDVSPSLDYEITTEKTYTLVRADGGKPLSNVSGLTVQVLDINDNAPLIVST

PFQATVLESLVGLVTLVQALDADAGDNARLETRLAGVGHDPPTTNGTWNISVA  
 ELDEBEVDFYSFGEARHDGTPTALTSASVSVTLVDVNNPFTQPEVTLREDA  
 VGTSVTVASVADDAHSVITVQITSGNTRRPSITSOSGGVSLPDLKLEERYV  
 LATVADGTRODPAQIVVNTDANTHRPPOSHTYVNNRBPACTVYLSANDED  
 TGNMATTYRMEHSIPQRIADATGAVTTQALDYEDQVSYTLATIRONGLEFQKSDT  
 TYLEIIVNDVNDAPQLRDSYQGSVEDEPPTSVLQISADRDGLNGRVEFYOG  
 GDGDDDFIVESTSGIVRTLRLDRNRVATYVRAVAKMGPATPMEVTVLDV  
 NDNPVPEODEPVPVENSPIGLAVARVATPDDEGTQAQVQIVEGNIPEVQLD  
 IFSGELTALVDLDEDRPEVYLVIQATSAPVSRATVRVRLDRNPNPLVNGEILP  
 NNYVTRSSSPGGAIGRVPAPNDPISDLSTVFERGNRLSVLNAISQIKLALP  
 DNNRPIELNINSVVSQSVHSVTAQCALRTIITDEMLTSHITRLKEDMPEERLSTPL  
 GLTQVAAATLAPDPAHVFNQDPTDAGHILNVSLSVQPPGPGRGAPVEYFED  
 LQRLVNLNLSLTAISAORVLPDNDICLRBPCENMRCVSPAPPSAPLSAVL  
 FRPIHVGQGLRCRCPGFTDCEYVDICSPGPHGCRREGVYICLSPASSTV  
 HECSEVARSGRCTPGVCKNGKTCVNLVGGFKDCPSGDEKREYCOVTTNRSPASHFI  
 TPRGLRORPHFTLASPTATKERDGLNMRFNKXDPVALVIOBONOLTPSAGST  
 TTVSPVPGVSPQCHTVOYLKYNKPLAQTCGLPGPSBOKAVVTVGDCDTVALR  
 PGLSVLNCISCAQGTQGSKSLDTGLPLAGVDPDPSFPVRMQRVQVGCMMNLVD  
 SRHIDMADFLANNQVPGCPAKNVCDSNCHNGTCVQNMDFSCBPCYFGKSCA  
 OEMANPOHFTGSSLVAMHGLSLPISQPMYSLMFRROADGVLLQAITRGRSTTLQ  
 REGHVMLSYEGTLOASLSLBEPGRANDGMHQAOLALAGSGRHAIIISPDYGOORA  
 BGNLGRHGLHLSNITVGGIPGAGVARGPRGICQVVRSPTPGSGVNSLDSHRS  
 INVQCSLPDPCDSNPCCPANSYCSNDMSYSCSCPRTYGDNCTVCLNPEHOSV  
 CTRKSPAHGTYCECPNNTLGPYCETRLIDPCPRGMWGHPTCGPCNDVSKGPDVDCN  
 KTRSGECHKENHRRPPSPCLDCYPTGSLRVCDPBDQCPCKRQVYGRQCDRC  
 NPEAVTYTNGCEVNVDSCPRAIEAGIWPETRGALPAAAPCPGYSOTAVRHCDBRG  
 WLPENFNCTSIYFSELKCPARELORNESGLDSRGOOLALRNANTOHTAGYFSDV  
 KVAYOLATRLAHSTORGRGLSATQDHYETTELTVGSLALDTRKRMELILOQEG  
 GRAMLOHREAVASALAKONRHITYLSPITVTNITYISVRLDKGNPAKAPRYTAL  
 RGSQPDLETTVILPESVRETPPVVRPARGAQBEPBLARQORHPBLISQBEAVAS  
 VLIYRTLAGLPHNDPDKSLNVPKRPINTVVASIYHDEBLPLRLDRKPVTVQF  
 RLLETERETPCVFNWNSILVSGTQMSARGCVFVRSHSVSCCNMTGSAVMD  
 VSRRENGELPLKTLTYVALGTVALILHFFPLTLRLIRSGOHLRNLTRVALGDA  
 OLVTLEINGNOALPACTVIALHFFLYICTPSMALLBMLHRYALITERDVRGEMR  
 FYRMGMPARITGLANGLDPRGYNPPCMISTYDTLINSAGVYARAVSKSVLY  
 ILAARSCAQRQGEKQFVSGLOSPFALHLSATFWLALSIVSUTLLEPLAT  
 CNCTIGPFLPLSTVLSKEVRKALKASCSPDPALTKSLTSTSVSPSYADGR  
 LVQPYGDSAGSLSTSRSGKSPSYIPLRLRESALNPGCPGADPSLFPLEQGD  
 OHDPYDSDSLDEDDQSGSYASTHSDBEEREBEEREBEAPAPGQMDSLGAGQ  
 RLPFSTPKDQGPAGKAPWPDGRTTARGTGAAGPBEERLNGALRREGSLGLP  
 GSSAOKHGIKKKCLPTISEKSLRLPLEOCTGSGSSASBSRGCPRRPPRPO  
 SLOBQNGWPMIIMSTKATVDEDSGSEFLFPNPLH"

BASE COUNT 4543 a 7241 c 7458 g 5128 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 length: 24370  
 Score: 1103.00 Matches: 1103  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 37.74% Indels: 0  
 DB: 6 Gaps: 0

US-09-916-849a-3 (1-2923) x AK646739 (1-24370)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 DB 201 ATGCGAGCCCGGCGACCGGCGTCCCTCCCAACCGCGCGCGCGTGTGCTGCTG 260  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysValGserLeu 40  
 DB 261 TTCTGTGTGTGTGCGCGCGCACTATTGGAGAGCAAGTGGGGCCCTGTCTCTTTC 320  
 QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyThrLeuLeuCysProSer 60  
 DB 321 TTAGCGGCGGAGAGAGGCTCTTCGCGGGGCTTCGCGCCCAATGGGCTGTCTTCATCC 380  
 QY 61 SerAlaSerAsnLeuThrLeuLeuYThrSerArgCysValGAspAlaGlyThrGluLeuThr 80  
 DB 381 TCAAGCGTGAACCTGTGCTCTACACCAAGCGCTGAGGATCGGGCACTGAGCTGACT 440  
 QY 81 G1YHISLeuValProHISHisAspGlyLeuValGlyValThrCysProGluSerGlnaHis 100

DB 441 GGCACCTGGTATCCCAACCAAGATGCGTGAAGGTTTGGTGTCCAGAAATCCAGAGCCCAT 500  
 QY 101 ILeProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
 DB 501 ATTCCCTTACCAACAGCTCTCTAAGAGCTGCCCCCTGAGCTGTGCGCTCTTGAGCATGGA 560  
 QY 121 G1YHISLeuSerProGluGlyIleLeuThrLeuProGluGluHISProCysLeuVala 140  
 DB 561 GGCACCTTTTCCCAACAGGAGGATCACTGCGCCGAGAGACACCGGTCTTAAGAGCT 620  
 QY 141 ProHISLeuArgCysGlnSerCysIleLeuAlaGlnAlaProGlyLeuValaGlyIle 160  
 DB 621 CCACCGCTCAAGATGCGACGCTCGCAAGCTGGGACAGGCCCCCGGCTCAGGGCAGGGGA 680  
 QY 161 ArgSerProGluGluSerLeuGlyIleValArgGlyValArgAsnValAsnThrAlaProGln 180  
 DB 681 AGGTCAACAGAAAGATCCCTGAGTGGGCTCGAAAGAAAGAAATGAATACACCCCCCA 740  
 QY 181 PheGlnProProSerTyrglnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 741 TTCAGCCCCCAAGCTACAGGCCACAGTGGCGGAGAACACAGCCAGAGGACCCCTGTT 800  
 QY 201 AlaserLeuArgAlaIleAspProAspGlyIleGlyValaGlyValGluGlyThrThr 220  
 DB 801 GCATCCCTGAGGCGCATGCAACCCGACAGAGGTGAGGAGGTGCACTGGAGTACACCATG 860  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyVala 240  
 DB 861 GATGCCCTTTTGAATAGCCGCTCCAAACAGATTCTTCTCTGAGCCACGTCACTGTGCA 920  
 QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrIleYserThrHisValaPheArgValThr 260  
 DB 921 GTAAACACAGCGGAGAGGTGATGTAAGACCAAGAGACCAACCTCTTACAGGCTACAG 980  
 QY 261 AlaglnAspHisGlyIleCysProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 981 GGGCAGAGAACCAAGGATGCCCGAGAAAGTCCCTGCGTCACTACCAATCTTGATTACT 1040  
 QY 281 AspThrAsnAspHisAspProValPheGluGluGluGluIleValYserLeuArgGlu 300  
 DB 1041 GACACCAATGACCAATGACCTCTGTCTTCAAGCAGCAGAGTACAAAGAGAGCTCAGGAG 1100  
 QY 301 AsnLeuGluValaGlyTyrgluValLeuThrValaArgAlaThrAspGlyValAspAlaProPro 320  
 DB 1101 AACCTGAGGTTGGCTATGAGGTGCTCACTGTCAGGCGCAACGAGTGTGATCCCTCC 1160  
 QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGluGlySerGlyIleYserProSerGluValaPhe 340  
 DB 1161 AATGCCAATATTTCTGTACCGCTGTGAGAGGGGTCTGGGGGCGAGCCCTCTGAAGCTTT 1220  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 DB 1221 GAGATGACCTCGCTCTGCGGTGATCCGAAACCTGTGCGCTGTGATGCGGAAGAAGT 1280  
 QY 361 GluSerTyrglnLeuThrValaGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 DB 1281 GAATCTTACCACTGACGTAGAGCAAGTGAACAGGGTCGGAGACCGGGGCTCTCGAGAT 1340  
 QY 381 ThrThrAlaAlaValaPheLeuSerValaGluAspAspAsnAspAlaProGlnPheSer 400  
 DB 1341 ACCACAGCGCGTGTTCCTTCTGTGTGAGAGVGAACAAATGAATAGCCCCCAAGTTTAT 1400  
 QY 401 GluTyArgTyValaValaGluValaArgGluAspValaThrProGlyValaProValLeuArg 420  
 DB 1401 GAGAAGCGGTATGTGTCCAGGTGAGGAGATGAGTCCACAGGGGCCCAATATCTCGGA 1460  
 QY 421 ValThrAlaSerAspArgAspIleGlySerAsnAlaValaHisIleYserIleMetSer 440  
 DB 1461 GTCAACAGCTTCGATCGAGACCAAGGAGAGCAATGCCGTGTGTGACATATGACTGAGAT 1520  
 QY 441 G1YAsnAlaArgGlyIlePheTyThrLeuAspAlaGluThrGlyValaLeuAspValaIleSer 460  
 DB 1521 GGCATGCTCGGGAGACATTTTATCTGATGCCAGACTGAGAGCTGTGATGTGTGAGC 1580



QY	461	ProLeuAspTyrGluThrThryrsglyTyrThrLeuArgValArgAlaGlnAspGlyGly	480
DB	1581	CCCTTGACTATAGACGACCAAGAGTACACCTACGGGTGAGACACAGATGTGGC	1640
QY	481	ArgProProLeuSerSerValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp	500
DB	1641	CGTCCCACTCTATATGCTCTGCTGGTGAAGTACAGGTCTCTGATATCAACGAC	1700
QY	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu	520
DB	1701	AATGCCCACTCTCTCTACGACACCTTTCAGGCTACTGCTCTGAGAGTGTCCCTTA	1760
QY	521	GlyTyrLeuValLeuHileValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
DB	1761	GGCTACTGCTCTCTCATGCTCCAGGCTATCGAGCTGATGCTGTGACAAATGCCGCTG	1820
QY	541	GlyTyrArgLeuAlaGlyValGlyHileAspPheProPheThrIleAsnAsnGlyThrGly	560
DB	1821	GAATACCGCTTGTGTGGGTGGGACATGACTTCCCTTCAACATCAATGGCAAGGC	1880
QY	561	TyrIleSerValAlaAlaGlnLeuAspArgGluGluValAspPheTyrSerPheGlyVal	580
DB	1881	TGGATCTCTGTGGCTCTGAACCTGACCGGAGAAATTGATTTCTACAGCTTTGGAGTA	1940
QY	581	GluAlaArgAspHileGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
DB	1941	GAAGCTCGAGCAATGGCACTCCAGCACTCACTGCTCGGCAAGTGTCAAGCTGACTGC	2000
QY	601	LeuAspValAspAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620
DB	2001	CTGGAGTCTCAACGACAAACATCCACTTACCCCAACAGATACCACTGTGGCTCAT	2060
QY	621	GluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640
DB	2061	GAGATGCAAGCTGTGGGACCAAGCTGGTGAAGGTGACGTGTGACCTGATGCTCAT	2120
QY	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660
DB	2121	AGTGCATCACTACCAATCAACAGTGGCAATACCTGAAACCGCTTCTCATCAACCAAC	2180
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln	680
DB	2181	CAAAAGTGGTGGTGGCTGTATCCCTGCTGCCCTGCACTGAGTACAAATTTAAGGGCAG	2240
QY	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700
DB	2241	TATGTGTGGCTGTATCCGCTCCGATGGCACTCGGACGACACGGCACAGATTGGTG	2300
QY	701	AsnValIleThrAspAlaAsnThrHileArgProValPheGlnSerSerHileTyrThrValAsn	720
DB	2301	AATGCACCGACGCAACCACTGCTGTCTTCAAGACTCCCACTATACATGAT	2360
QY	721	ValAsnGluAspArgProAlaGlyThrThrValIleLeuIleSerAlaThrAspGluAsp	740
DB	2361	GTTAATGAGGACCGGCGGCAAGCAACGAGTGTGATCAGGACCGACCGATATAGAC	2420
QY	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGluPheArgIle	760
DB	2421	ACAGGTGAGAGTCCCGCATCACTTCTCATGAGAGACAGATCCCCAGTTCCGCAATC	2480
QY	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer	780
DB	2481	GATGACGACACGGGGGCTGTCAACACCGAGCTGAGTGAATGAGAGACCAAGTGTCT	2540
QY	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrTyr	800
DB	2541	TACACCTCTGGCCATTAAGTCTCTGGGACATGGCAATTTCCCAAAATCCGACCACTTAC	2600
QY	801	LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820
DB	2601	CTGGAGATCTCTGTGTAACGACTGAATGCAATGCCCTCTGACTTCTCTGAGACTCTTAC	2660
QY	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
DB	2661	CAGGCAAGTGTATAGAGATGTGCCACCTTCACTAGGTCTCTCAATCTCACCACT	2720
QY	841	AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyGlyAspAspGly	860
DB	2721	GATCTGTATCTGGAATTAATGGCAGGGTCTTCTACCTTCAAGAGGCGACATGGA	2780
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880
DB	2781	GAGGAGACTTATATGTGATGTCACGTCAGGCAATGTGGCAACCTTACGAGAGCTGGAT	2840
QY	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro	900
DB	2841	CGAGAGAAAGTGGCCCAATATGTCTTGGGGGATATGAGTGAACAAGGAGATGCCCA	2900
QY	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal	920
DB	2901	GCCCGCACCTATGTGAAGTGAAGTCACTGTGTGTGATGTGAATGACAAATCCCTGTTC	2960
QY	921	PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal	940
DB	2961	TTTGACGAGATAGTATGATGTGTGTGTGAAGAAGAACGCCATTTGGCTAGCCGTG	3020
QY	941	AlaArgValIleThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle	960
DB	3021	GCCCGGTCACAGGCACTGACCCGATGAAGGACCAATGGCCAGATATATGACAGATT	3080
QY	961	ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla	980
DB	3081	GTGGAGGGCAACATCCCTGAGGTCTTTCAGCTGGACATCTTCTCCGGGAGACTGCACACC	3140
QY	981	LeuValAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThrSer	1000
DB	3141	CTGGTACCTTACACTACAGACACCGGCTGATGACGTCTGTGTATCAAGGCCACGTCA	3200
QY	1001	AlaProLeuValSerArgAlaThrValHileValArgLeuLeuAspArgAsnAspAsnPro	1020
DB	3201	GCTCCTGTGTGAGCCGGGCTACAGTCCAGCTCCGCTCTTGAACCGAATGACAAACCA	3260
QY	1021	ProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValIleThrAsnArgSerSerSer	1040
DB	3261	CCAGTGTGGGCAACTTGAATCTTTTCAACAACATATGCACCAATCGTCAAGACGC	3320
QY	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
DB	3321	TTCCCTGGGGGTGCATTTGGCCGAGTACTGCGCATGACCTGTATCTCAGATATGTTG	3380
QY	1061	ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080
DB	3381	ACTTACAGCTTTGAGCGGGGAATGAATCACTGAGTCTGTCTCAATGCTCCACGGGT	3440
QY	1081	GluLeuIleLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal	1100
DB	3441	GAGCTGAAGCTTACACCGGCACTGAGACAAACCGGCTCTGAGAGGCACTATGAGCGTG	3500
QY	1101	LeuValSer 1103	
DB	3501	CTGGTGTCA 3509	
RESULT 8			
AB065955			
LOCUS			
DEFINITION			
Homo sapiens gene for seven transmembrane helix receptor, complete			
cde, isolate:CBRC7M_518.			
ACCESSION			
AB065955			
KEYWORDS			
AB065955.1 GI:21929187			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
1			

AUTHORS Suwa, M., Sato, T., Okouchi, I., Arita, M., Futami, K., Matsumoto, S., Teutsumi, S., Aducci, H., Asai, K. and Akiyama, Y.  
TITLE Genome-wide discovery and analysis of human seven transmembrane helix receptor genes  
JOURNAL unpublished  
REFERENCE 2 (bases 1 to 24370)  
AUTHORS Suwa, M.  
JOURNAL Direct Submission  
TITLE Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan (E-mail: m-suwa@aist.go.jp, URL: http://www.cbrc.jp/, Tel: 81-3-3599-8080, Fax: 81-3-3599-8083)  
COMMENT This sequence is a seven transmembrane helix receptor candidate predicted from the whole human genome sequences using our automated system that contains programs of gene finding (GenesDecoder), sequence search, motif-domain assignment and transmembrane helix prediction.  
And the sequence is submitted by the collaborative project between [Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)] and [Genome Science Division, Research Center for Advanced Science and Technology (RCAST), University of Tokyo].  
FEATURES  
source location/Qualifiers  
1. 24370  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/isolate="CBRC7TM\_518"  
/db\_xref="taxon:9606"  
/chromosome="1"  
101n(201..3510,8553..9200,11163..11385,11634..11749,11929..12017,12408..12565,12927..13090,13263..13388,13732..13898,14199..14486,14575..14749,14989..15130,15312..15432,15856..16039,16226..16327,17669..17763,17972..18181,18702..18885,19001..19129,19231..19411,19545..19712,19815..19941,20054..20254,20517..20721,21048..21160,21338..21417,21507..21633,21721..21843,22399..22526,22761..22849,22954..23148,23287..23457,23557..23791,24143..24170)  
/codon\_start=1  
/evidence=not experimental  
/product="seven transmembrane helix receptor"  
/protein\_id="BAC06168.1"  
/db\_xref="GI:21929188"  
/translation="MRSFANGVLPPLPPLLLLLPPLLDQVPCRSLSRGKSSGACAPWGLCPSSANGLMLYSCRCDAGTELGHVPHDGLQVCPSEKAIPLPAPGCGWSCGLIGLGHLSPOGLTLPBHPCLKAPRLCOSCKLAOPABAGE RSPESLGGRRKKNVTPAPOPPOPSOATVPENOPACTGPAASRAIDPREGARELEY TMALPDSNSQPSLDPVGAATTAELDRKTSVTRAVTAQDGMPPRSALATLT ILVTDNHDVPEQOEKREKRENEVEGLVTRATGDAENPNNILYRLBSGG SESEVEIDPRGVIRTRGPVREKRENEVEGLVTRATGDAENPNNILYRLBSGG NDNAPOFSEKRYVQVREDEVTGAPVLRVYASDRKSNVAVYISMSGAGQFLD AOTGALDVSPLEVTREKYLTVRAQDGRPLSVNGSLVGVLDINDNAPIFST PRQATVLESPLGVLYLAVOADAIDAGDNARLEVRLAGVLDHPPTINNGTMSAA ELDRVDFPISGVYARADHETPLTASASVTVLVNDNNPFTPTETVRLNEBA VGVSVVSAVDADHSAVITVQITSNATNRFSTISQSGGLVSLPLDYKLERVY LATASDGTRODIAQIVVNTANTHRPVQSHSYVNVNEDPACTVTLISATD TGENARITVEMEDSIPOFRIDADGATTAELDEYDQVSYTALITARPNGVLOKSDT TYLEILVNDVNDAPOLRDSYQSGVYEDVPPTVLSATDRSGALGVYTVQV GGDGDDPVTSESTGIVRTLRDRENVAVPVLAVYAKMCPBAGDAPPEVTVYV DNDNPVPEQDEPVTVEBNSPGLAVARTATPDDEGTNAQITVQVEKNIPVFD ID IFSEGLTALVDLIDYEDRPEVLAIVQISALVSRATVRLDRNPNVLAQFLIF NNVYTRSSSPGAGIAGRPAHPDISDLSYSPRGNELSLVLNLSGELSLAL DNRRPLAIVLSVSDGVSHVTAQCALRVITIDEMLTSHITPDRGSEPERLSL GLFLOVAATLATPDRHVAVVNOPLDAGGHIILNLSVQOPRPGGAPPPLPDED LQRLVILNRSILTAISORVLPDDNICALBPCNMCVSRVAPSPASAPLASSVTV FRPIHVGLRCRCPGRTGDCVETVDLCYSPGCBHRCRSGEGYICIDQYGT EKVGVASRGCTPGVCKNGCTCVNLVSGFKDCPSGDEKRYCOVITRSFPAHPTI TFQGLRQRFITLALSFAIKERDGLILYNREKIDPVALVIAOVODUTSFGABST TVSPVPGVSDGOMTVQLKYNKRFLVQGLPOGSPQKVAIVTVVGDCTGVLR FGSVLNDSYCAAGTQSGSKSLDITGPIILGVPDLPSEFPRMRCQVYCKMNLVD SRIIDMADFIAANGTVGCGPAKNCVCSGNGTCCNGTCCNMWDMPSECPGFGSKCA QEMANPQHFIAGSSLVAMHGLSLPISQPWYLSMFRTRQADGVLLQAITTGRSTITLOL

REGHVLVEGTGLQASSLRLEPRGRANDGMWHAOLALASGSPGHAILSPYGOORA  
BENLDPRLHGLHLSNITVGGIIPGAGVAGRGGLQGRVSDTREGNSLBSGES  
INVEGCSLPDPSNPNCPANSYCSNDMSYSCSDPGYGNCTNVDINPCHESV  
CTKRSASHGTCCECPNYLPYCESTRIDQEPKRWKHPPTGCPKCVSKSPFDCN  
KTSRCHCKENHTRPPSPPTCLCDCTYFSLRVCDEPDGCPKPGVIGRCQRC  
NPEAVTNGCEVNVDSCPRALIEAGIWPRTFRFLPAAAPCKSGFVAVRQDCBRD  
MLPMLFNCTSLTSEBKGFARLQNRNMSGDLSGOOLALRNATHTVAGPESDV  
KVALATRLALHESSTORGFGLSATODVHFTNLSVGSALDITANKHWEILQOTEG  
GTAMLOHRYVASALANMRHTYSLPTVPTNPIVSVLRDKNPAVAGAKLRREAL  
RGEQDPLLETVYILDESVPRTPVPVPRAGCEVPEVRLARORRHPPLSGEYAS  
VITITLGLLPHNDPDKRSIRVKEPITINTPVVISVHDELLPRALDQVTVQ  
VLLREERTKPLCYFNWNSILVSGTGSNARCEVFNRESVSCQCNMTSPALMD  
VSRRENGBILPLKTLTYVALGVTLALITLPEPLTLRLNNOHGIKRNLPALGLA  
OLVFLGINOMDLPACTVIALILHFLYCTFSMALBLALHLYALTEVADVNTGEMR  
FYVMGQVPAVITGLAVGLDBEGVNDPFCMLSYDYLIMSFAQVPAVAVMSVLYL  
TLAASCAARQGRPKGVPVGLQSPSAVILIASVTLALISVNSDTLPHYLPAT  
CNCTGPFITFISVTVLSKVRKALGLAKSRSPDPALTTSTLTSINCPSPYADGR  
LYQPYGDSAGSLHSTRSRGKQPSYIPLLEBSALNFGQPGGLADPGSLBQDQ  
OHPDPTDSDSLDELDQSGYASVYSHSDSEBEBEBEBEAPBQDMSLGLGAE  
RLHPSSTPDGPGPKAPWPDFTGTAKSSGNGAPBRLNENDALRSGLPLP  
GSSAQPHGILKKKCLPTISEKSLRLPLEQCTGSSSASBSRGPPRPPRQ  
SLQRLNCTVMPITMSIKAGTVDESDSSBFLFPLNH"  
BASE COUNT 4543 a 7241 c 7458 g 5128 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 24370  
Score: 1103.00 Matches: 1103  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 37.74% Indels: 0  
DB: 9 Gaps: 0  
US-09-916-849a-3 (1-22923) x AB065955 (1-24370)  
QY 1 MetAAGSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
DB 201 ATGCGAGGCCCGGCGACCGGCGTCCCGCCAGCGCGCCCGCGCTGTGTCGCG 260  
QY 21 LeuLeuLeuLeuLeuProProProLeuGlyAaPAGLInaIGLYProCYaSerSerLeu 40  
DB 261 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320  
QY 41 GlySerAaGlyAaGlySerSerGlyAaCYaAlaProMetGlyTTPLeuCYaProSer 60  
DB 321 GGGTCAAGGAGGAGGAGGCTTCCGAGGCGCTGCGCCCAATGGCTGCTGCTCATCC 380  
QY 61 SerAlaSerAaLeuThrLeuTyThrSerAaGlyAaGaaAaAlaGlyThrGluLeuThr 80  
DB 381 TCAGGCTGCAACCTGTGCTTCAACACAGCCGCTCAGAGATGCGGCACTAGCTGACT 440  
QY 81 GlyHLSLeuValProHLSHLSAaPGLYLeuAaGValTTPCYaProGluSerGluAaHLS 100  
DB 441 GGCACACTGATGCCACACACATGGCTGAGGCTTGGTGTCCAGAAATCCAGGCCCAT 500  
QY 101 1LeProLeuProProAlaProGluGlyCYaProTTPSerCYaAaGLeuGly11egly 120  
DB 501 ATTCCTTCAACACAGCTCTTAAGAGGCTGCCCTGAGGCTGTGCTCTCTGAGCATTTGA 560  
QY 121 GlyHLSLeuSerProGluGlyLeuLeuThrLeuProGluGluHLSProCYaLeuAaHLS 140  
DB 561 GGCACACTTTCCTCCACAGGAGGCAAGCTCACATGCCCGAGAGACACCCGCTTAAAGGCT 620  
QY 141 ProAaGLeuAaCYaGluSerCYaHLSLeuAaGlnAaProGlyLeuAaGlaGlyGlu 160  
DB 621 CCAAGGCTCAATGCAAGTCCGCAAGCTGCAAGGCGCCCGGAGTCAAGGCAAGGGA 680  
QY 161 AaGserProGluGluSerLeuGlyGlyAaGlyAaGaaAaValAaThrAaProGlu 180  
DB 681 AGGTCAACCAAGAAAGTCCCTGGGTGGGCTCGAAAGAAAGTAAATACACCCCCAG 740  
QY 181 PheGlnProProSerTyGlnAlaThrValProGluAaGlnProAlaGlyThrProVal 200

Dh	741	TTTCAGAGCCCCCAGCTACCAAGCCACAGTGGCCGGAGAAACCAAGCCAGACAGGACCCCTGTT	800
Qy	201	AlAserLeuAArgAlaIleAspProAspGluGlyGlnAlaGlyArgLeuGluIlyrThrMet	220
Dh	801	GCATCCCTGAGAGGCCATGAGACCCGGACGAGGGTGAAGGAGGTGCATGGATGACCAATG	860
Qy	221	AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla	240
Dh	861	GATGGCCCTCTTGGATAGCCGCTCCAAACAAGTTCCTTCCCTGGAGCCCAAGTCACTGGTAC	920
Qy	241	ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr	260
Dh	921	GTAACCAACAGCCGAGAGAGCTGGATGCTGAGAACCAAGACCAACAGTCTTCAAGGGTCAAG	980
Qy	261	AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr	280
Dh	981	GCGCAGAGACCAAGGATGCCCGAGAGAGTGGCCCTGAGTCACTACCTACCATCTGGTTACT	1040
Qy	281	AspThrAspAspHisAspProValPheGluGlnGlnGluIlyrLysGluSerLeuArgGlu	300
Dh	1041	GACACCAATGACCAATGACCCCTGTGTTCCAGACGACGAGGTCAAGAGAGAGCTCAGGAG	1100
Qy	301	AsnLeuGluValGlyIlyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro	320
Dh	1101	AACCTGGAGGTGGCTATGAGGTGCTCACTGTCAAGAGGACACAGAGTGAAGGCCCTCC	1160
Qy	321	AsnAlaAsnIleLeuIlyrArgLeuLeuGluGlySerGlyLysProSerGluValPhe	340
Dh	1161	AATGCCAATATTCTGACCGCTGCTGAGAGGGGTCTGGGGGCGAGCCCTCGAAGTCTTT	1220
Qy	341	GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal	360
Dh	1221	GAGATCGACCTCGCTCTGGGGGTGATCCGAACCCGTGGCCCTGTGTGATCGGAGAGAGGTG	1280
Qy	361	GluSerIlyrGlnLeuThrValGlnAlaSerAspGlnGlyValArgAspProGlyProArgSer	380
Dh	1281	GAATCTCAACAGCTGACCGGTAGAGGCAAGTACCAAGGGTCCGAGACCCGGGTCTCCGAGT	1340
Qy	381	ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer	400
Dh	1341	AACCAAGCCGCTGTTTCTTCTCTGTGAGAGATGACAAATGATATATGCCCCCAATTAGT	1400
Qy	401	GluLysArgIlyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg	420
Dh	1401	GAGAAAGCCCTATGTGTGTCAGAGTGAGGAGATGATCACTCAAGGGGCCCCAGTACTCGA	1460
Qy	421	ValThrAlaSerAspArgAspLysArgLysSerAsnAlaValAlaIlyrSerIleMetSer	440
Dh	1461	GTCAAGAGCTCGAGTACGACAGCAAGGGAGCAATGCCGTGGTCACTATAGATCATAGAT	1520
Qy	441	GlyAsnAlaArgGlyGlnPheIlyrLeuAspAlaGlnThrGlyValAlaLeuAspValValSer	460
Dh	1521	GGCAATGCTCGGGGACAGTTTATCTGGATGCCAGACTGAGAGCTCTGGATGTGTATGC	1580
Qy	461	ProLeuAspIlyrGluThrThrIlyrGluIlyrThrLeuArgValArgAlaGlnAspGlyGly	480
Dh	1581	CCTCTTGACTATGAGACGACCAAGAGTACACCTTACGGGTGCCAGACAGAGATGCTGTGC	1640
Qy	481	ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp	500
Dh	1641	CGTCCCCACACTCTCAATGTCTCTGGCTTGGTGAACAGTACAGGTCTCGAATATCAAGAC	1700
Qy	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluIlyrValProLeu	520
Dh	1701	AATGCCCCCACTTCTGTCAGACACCCCTTTCCAGGCTACTGTCTCGAAGAGTCTCCCTTA	1760
Qy	521	GlyIlyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Dh	1761	GGCTAACCTGGTTCTTCATGTCCAAGCTATTCAGCGTATGCTGGTGAACAATGCCGCTG	1820
Qy	541	GluIlyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560
Dh	1821	GAATACCGCCTTGTCTGGGTGGAGCATACCTTCCCTTCAACATCAACAATGGCAGAGGC	1880

QY	561	TPPILESerValAlaIaGIuLeuAspArgIuGIuValAspPheTYrSerPheGIuVal	580
Db	1881	TGAATCTCTGGCGCTGCTGAACTGGAACGGGAGGAAGTGAATTCATCAAGCTTGGGGTA	1940
QY	581	GIuAlaArgAspPheIaGIuThrProAlaLeuThraIaSerAlaSerValSerValThrVal	600
Db	1941	GAAGCTCGAAGCAATGGACTCCAGCACTCACTGCTGGCAAGTGTCAAGGTGACTGTTC	2000
QY	601	LeuAspValaAspAspAspProThrPheThrGIuProGIuTYrThrValArgLeuAsn	620
Db	2001	CTGAGTGTGCACGACACAAATCCAACTTTCACCAACAGAGTACACAGTGGCTCAAT	2060
QY	621	GIuAspAlaAlaValGIuThrSerValValThrValSerAlaValaAspArgAspAlaHis	640
Db	2061	GAGAGTCAGCTGTGGGCAACAGGTGTGTACAGGTGTCACTGTGGACCGTGAAGCTCAT	2120
QY	641	SerValIleThrTYrGIuIleThrSerGIuAsnThraArgAsnArgPheSerIleThrSer	660
Db	2121	AGTGCATCACTTACCAAGATCACCAAGTGGCAATCTGAAACCGCTTCTCATCACAGC	2180
QY	661	GIuSerGIuGIuGIuValaValSerLeuAlaLeuProLeuAspTYrIleGIuGIuArgGIu	680
Db	2181	CAAGTGTGTGTGGCTGTGATCCCTTGGCTCCCTGGCACTGACATCAAACTTGACGGCAG	2240
QY	681	TYrValLeuAlaValThraIaSerAspGIuThraArgGIuAspThraIaGIuIleValIaI	700
Db	2241	TATGTGTGGCTGTACCGCTCCGATGGCACTGGCAAGACACGGACAGATTTGTGCTG	2300
QY	701	AsnValThraAspAlaAsnThraAspProValPheGIuSerSerHisTYrThrValAsn	720
Db	2301	AATGTCAACCGACGGCAACCCCAATCGTCTGTTCATGAGCTCCCACTTACAGTGAAT	2360
QY	721	ValaAsnGIuAspArgProAlaGIuThrThnValIleuIleSerAlaTrpAspGIuAsp	740
Db	2361	GTTATATAGACACCGCGCGACAGGACCAAGGTGTCTGATCAAGCCGACAGATAGACAC	2420
QY	741	ThrGIuGIuAsnAlaArgIleThrTYrPheMeGIuAspSerIleProGIuPheArgIle	760
Db	2421	ACAGGTGAGATGCCCAATCACTTATGAGGAGACAGCAATCCCCAGTTCCGATC	2480
QY	761	AspAlaAspThrArgIuAlaValThrThrGIuAlaGIuLeuAspTYrGIuAspGIuValSer	780
Db	2481	GATGAGACACGGGGGCTGTCAACCAACGAGCTGAGCTGAGCTTGAAGACCAAGTGTCT	2540
QY	781	TYrThrLeuAlaIleThraIaArgAspAspArgIleProGIuIlySerAspThrThrTYr	800
Db	2541	TACACCCCTGGCAATTACTGCTCGGACAAATGGCAATTCACCAAGTCCGACACCACTAC	2600
QY	801	LeuGIuIleLeuValaAspValaAspAspAlaProGIuPheLeuArgAspSerTYr	820
Db	2601	CTGAGATCTCCGTGGACGACAGGAAAGAAAGAACCCCTCAAGTTCGCGACAGACTCTAC	2660
QY	821	GIuGIuSerValTYrGIuAspValProProPheThrSerValLeuGIuIleSerAlaThr	840
Db	2661	CAGGGCAGGTGTATAGATGAGATGTGCCACCTTCACTAGCTCTCGACAGACTCAAGCACT	2720
QY	841	AspArgAspSerGIuLeuAsnGIuValArgValPheTYrThrPheGIuGIuIlyAspAspGIu	860
Db	2721	GATGTATCTTGGACTTAATGGCAGGAGGTCTTCACTTCCAAAGGAGCGACAGATGGA	2780
QY	861	AspGIuAspPheIleValGIuSerThrSerGIuIleValaArgThrLeuArgValLeuAsp	880
Db	2781	GACGGTACTTATATGTGAGTCAAGCTCAAGGCAATGTGGAAAGCGCTACGAGAGCTGAT	2840
QY	881	ArgGIuAsnValaIaGIuTYrValLeuArgAlaTYrAlaValaAspArgIleProPro	900
Db	2841	CGAAGACAGTGGCCAGTATGTCTTGGCGGCAATATGCAATGAGCAAGGGAGTGGCCCA	2900
QY	901	AlaArgThrProMeGIuValaThrValThrValLeuAspValaAspAspProProVal	920
Db	2901	GCCCGCACCTTATGAGATGACAGTCACTGTGTGTGATGTGAATGACAAATCCCCCTGTG	2960

```

Qy 921 PheGluGlnAspGluPheAspValPheValGluGlnAsnSerProIleGlyLeuAlaVal 940
Db 2961 TTTGACACGAGATAGTTGATGTTGTGAGAGAGACAGCCCATTTGGGCTAGCCGGT 3020
Qy 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db 3021 GCCCGGGTCCACGACCTGACCCCGAGTAAAGGACCAATGCCAATATGATGACGAAAT 3080
Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 3081 GTGAGAGGGAACATCCCTGAGGCTTTCACTGAGCATCTTCTCCGGGAGCGTGAACGCC 3140
Qy 981 LeuValAspLeuAspTyrGluAspA-gProGluTyrValLeuValIleGlnAlaThrSer 1000
Db 3141 CTGTGAGACTTAACTACAGAGGACCCGCTGAGTACGCTCTGTGATCCAGGCGACGTCA 3200
Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020
Db 3201 GCTCTCTGTGAGCGGCGGTACAGTCCAGTCCCTCTTACCGCAATGACCAACCA 3260
Qy 1021 ProValIleGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db 3261 CCAAGTCTGGGCAACTTTGAGATCCTTTCAACAATATGTCCATGCTCAAGCAAC 3320
Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3321 TTCCCTGGGGGTGTCATGGCCGACGATACCTGCCCATGACCTGATATCTCAGATAGTCTG 3380
Qy 1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db 3381 ACTTACAGCTTTAGGGGGGAAATGAACTGAGCTGCTGCTCAATGCTTCCACAGGAT 3440
Qy 1081 GluLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db 3441 GAGCTAGAGTAAACCGCGACATGACAAACAACGCGCTCTGAGGCGCATCAAGACCGTG 3500
Qy 1101 LeuValSer 1103
Db 3501 CTGGTGTCA 3509

```

```

RESULT 9
AL390252 169241 bp DNA linear PRI 08-JUN-2001
LOCUS Human DNA sequence from clone RP11-29704 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL390252
VERSION AL390252.9 GI:14346077
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169241)
Thomas D.
Direct Submission
Submitted (07-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requesters: clonerequest@sanger.ac.uk

```

## COMMENT

On Jun 11, 2001 this sequence version replaced gi:13273808. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given

## FEATURES

## source

In the feature table with their source databases: Emi, EMBL, Sw, SWISSPROT, Tr, TREMBL, Wp, WORMBEP, Information on the WORMBEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormep](http://www.sanger.ac.uk/Projects/C_elegans/wormep) This sequence was generated from part of bacterial clone contig of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP11-29704 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

## VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-29704 The true left end of clone RP11-173K24 is at 98796 in this sequence. The true right end of clone RP11-352P4 is at 101180 in this sequence.

## Location/Qualifiers

```

1..169241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_id="RP11-29704"
/clone_1fb="RPCI-11.2"
1676..1775
/note="MIR repeat: matches 45. 152 of consensus"
2574..3410
/note="CpG island"
/evidence=not_experimental
2809..2936
/note="64 copies 2 mer gg 60% conserved"
2814..2943
/note="13 copies 10 mer ggcggggcgg 63% conserved"
2982..2987
/note="sequence from overlapping clones BA352P4 (
AL356389) and d6641D22 (AL157901). Assembly confirmed by
restriction digest."
3007..3116
/note="11 copies 10 mer gcccgggggc 63% conserved"
3364..3411
/note="16 copies 3 mer tgc 83% conserved"
9894..10207
/note="AluBg repeat: matches 1. 296 of consensus"
10332..10617
/note="AluIo repeat: matches 10. 299 of consensus"
11683..12328
/note="CpG island"
/evidence=not_experimental
12551..12727
/note="MIR repeat: matches 10. 197 of consensus"
13355..13442
/note="44 copies 2 mer gg 65% conserved"
15405..15448
/note="22 copies 2 mer gt 75% conserved"
17917..17995
/note="MIR repeat: matches 45. 129 of consensus"
19254..19293
/note="4 copies 10 mer aggcctgggg 95% conserved"
19651..19956
/note="AluBg repeat: matches 5. 309 of consensus"
19972..20270
/note="AluSc repeat: matches 1. 296 of consensus"
26166..26195
/note="10 copies 3 mer gag 90% conserved"
29612..29726
/note="FLAM C repeat: matches 1. 118 of consensus"
29778..30077
/note="AluXc repeat: matches 1. 298 of consensus"
30135..30428
/note="AluIb repeat: matches 1. 298 of consensus"
30454..30616
/note="MIR repeat: matches 98. 256 of consensus"
32159..32470
/note="AluSp repeat: matches 1. 312 of consensus"

```

repeat\_region 32481..32780  
/note="AluX repeat: matches 1..298 of consensus"  
repeat\_region 34670..34803  
/note="MIR repeat: matches 123..250 of consensus"  
misc\_feature 36339..36835  
/note="CpG island"  
/evidence=not\_experimental  
repeat\_region 37313..37937  
/note="MER41B repeat: matches 1..635 of consensus"  
repeat\_region 37938..38051  
/note="MER41B repeat: matches 103..213 of consensus"  
repeat\_region 38052..38351  
/note="AluY repeat: matches 1..300 of consensus"  
repeat\_region 38352..38753  
/note="MER41B repeat: matches 213..635 of consensus"  
repeat\_region 38987..39182  
/note="MER41B repeat: matches 1..171 of consensus"  
repeat\_region 39183..39482  
/note="AluX repeat: matches 1..300 of consensus"  
repeat\_region 39483..39504  
/note="MER41B repeat: matches 170..191 of consensus"  
repeat\_region 39537..39824  
/note="AluX repeat: matches 1..304 of consensus"  
repeat\_region 39850..40383  
/note="MER41A repeat: matches 1..554 of consensus"  
repeat\_region 40394..40483  
/note="MER41B repeat: matches 359..448 of consensus"  
repeat\_region 40532..40838  
/note="AluY repeat: matches 1..311 of consensus"  
repeat\_region 40956..41169  
/note="L1P81 repeat: matches 5927..6151 of consensus"  
repeat\_region 41170..41472  
/note="MER41B repeat: matches 433..745 of consensus"  
repeat\_region 41473..41771  
/note="AluY repeat: matches 2..300 of consensus"  
repeat\_region 41772..42011  
/note="MER41B repeat: matches 745..981 of consensus"  
repeat\_region 42399..42702  
/note="AluX repeat: matches 1..304 of consensus"  
repeat\_region 42948..43256  
/note="AluY repeat: matches 1..310 of consensus"  
repeat\_region 44024..44141  
/note="MIR repeat: matches 21..147 of consensus"  
repeat\_region 44187..44622  
/note="MIR repeat: matches 77..547 of consensus"  
repeat\_region 44790..45097  
/note="AluX repeat: matches 1..306 of consensus"  
repeat\_region 45098..45232  
/note="AluX repeat: matches 1..136 of consensus"  
repeat\_region 46712..47013  
/note="AluY repeat: matches 1..305 of consensus"  
repeat\_region 47032..47323  
/note="AluX repeat: matches 1..292 of consensus"  
repeat\_region 47669..47826  
/note="L2 repeat: matches 2545..2709 of consensus"  
repeat\_region 48017..48204  
/note="MIR repeat: matches 22..209 of consensus"  
repeat\_region 49692..49988  
/note="AluX repeat: matches 1..297 of consensus"  
repeat\_region 54709..55019  
/note="AluX repeat: matches 1..311 of consensus"  
repeat\_region 55261..55380  
/note="60 copies 2 mer gt 79% conserved"  
repeat\_region 55267..55386  
/note="12 copies 10 mer gtgtgtgtgt 79% conserved"  
repeat\_region 55386..55419  
/note="17 copies 2 mer gt 88% conserved"  
repeat\_region 55756..56029  
/note="L2 repeat: matches 46..328 of consensus"  
repeat\_region 56258..56327  
/note="MIR repeat: matches 333..388 of consensus"  
repeat\_region 56425..56915  
/note="L2 repeat: matches 330..834 of consensus"

repeat\_region 56926..57541  
/note="L2 repeat: matches 2097..2748 of consensus"  
repeat\_region 58447..58523  
/note="MIR repeat: matches 49..146 of consensus"  
repeat\_region 59152..59300  
/note="MIR repeat: matches 29..185 of consensus"  
repeat\_region 60773..61014  
/note="MIR repeat: matches 4..262 of consensus"  
repeat\_region 61115..61227  
/note="MIR repeat: matches 36..150 of consensus"  
repeat\_region 61443..61519  
/note="L1M1 repeat: matches 6234..6314 of consensus"  
repeat\_region 61956..62063  
/note="MER5A repeat: matches 35..141 of consensus"  
repeat\_region 63402..63437  
/note="L18 copies 2 mer ta 80% conserved"  
repeat\_region 68999..69310  
/note="AluB repeat: matches 2..306 of consensus"  
repeat\_region 69476..69781  
/note="AluX repeat: matches 1..292 of consensus"  
repeat\_region 70577..70718  
/note="MER3 repeat: matches 2..155 of consensus"  
repeat\_region 71406..71699  
Alignment Scores:  
Pred. No.: 0  
Score: 1103.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 37.74%  
DB: 9  
Gaps: 0  
US-09-916-849a-3 (1-2923) x AL390252 (1-169241)  
QY 1 MerArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
DB 3330 ATGGGAGCCCGGACCGGAGTCCCTCCCAAGCGCGCGCGCTGCTCTG 3389  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
DB 3390 TTGGTGTGCTGCGCGCGCGCGCGCACTTATGGAGACCAATGGGCGCTGCTCTG 3449  
QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTProLeuCysProSer 60  
DB 3450 GGGTCCAGGGGACGAGGCTCTGCGGGCGCTGCGCCCAATGGGCTGCTCTGCAATCC 3509  
QY 61 SerAlaSerAlaLeuThrLeuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80  
DB 3510 TCAAGCTGAACCTCTGCTCTTACACAGCGCTGACAGGATGGGCACTGAGCTGACT 3569  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100  
DB 3570 GGCACCTGTATCCCAACCAAGATGGCTTGAAGGTTTGGTGTCCAGATCCAGAGCCCAT 3629  
QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
DB 3630 ATTCCCTTACACACAGCTCTGAAAGGCTGCCCTGGAGCTGCTGCTCTGAGCATTTGA 3689  
QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGluGluHisProCysLeuValAla 140  
DB 3690 GGCACCTTTCCTCCACAGGCGAAGCTCACCTGCGAGGAGGACCCGTCTTAAAGGCT 3749  
QY 141 ProArgLeuArgCysGlnSerCysValLeuAlaGlnAlaProGlyLeuArgAlaGlyIle 160  
DB 3750 CCACGGCTCAGATGCCAGTCTCTCAAGTGTGACAGGCCGCCGAGGACAGGAGGA 3809  
QY 161 ArgSerProGluGluSerLeuGlyValArgValArgAsnValAlaThrAlaProGln 180  
DB 3810 AGGTCAACAAGATGCTCCGTGGGTGGCGCTCGAAGAAAGATATTAATACGCCCCAG 3869  
QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
DB 3870 TTCCAGCCCTCCAGCTTCCAGGCCACAGTCCGAGAACACAGCCAGGACCCCTGTT 3929

QY 201 AIASeLeuArgAlaIleaspProaPgiugIyGluAlaGlyArgLeuGluYrThrMet 220  
DB 3930 GGATCCCTGAGGGCCATCCACCCGGAGCGAGGGTGAAGCTGACCTGAGATACACCATG 3989  
QY 222 AAPAlaLeuPheAspSerArgSerAasnGlnphePheSerLeuAspProValThrGlyAla 240  
DB 3990 GATGCCCTCTTGATATGCGCTCCAAAGTTCTTCTCCCTGGACCACTGAGTCACTGGTGA 4049  
QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLySerThrHisValPheArgValThr 260  
DB 4050 GTAAACCAAGCGAGAGAGTGGATTCGTGAGACCAAGACACCACTCTTCAGGGTCAAG 4109  
QY 261 AIAgInaSPHieGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 4110 GCGCAGGACACAGCGCATGCCCCCAACAGATGCTGCTGCTACACTCACTCTGGTTACT 4169  
QY 281 AAPThrAsnAspHisAspProValPheGluGlnGlnGluYrThrLySerLeuAspArgGlu 300  
DB 4170 GACACCAATGACCATGACCTGTGTGAGCAGCAGAGATACAGAGAGCTCAGGGAG 4229  
QY 301 AsnLeuGluValGlyYrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 4230 AACCTGAGAGTGGCTGCTATGAGGTGCTCACTGTCAAGGCCACCGATGTGATGCCCTCC 4289  
QY 321 AsnAlaAsnIleLeuYrArgLeuLeuGluGlySerGlyYrSerProSerGluValPhe 340  
DB 4290 AATGCCAATATCTGATACCGCGCTGCTGGAGGGGTCTGGGGGACGCCCTCTGAAGCTTT 4349  
QY 341 GJuIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
DB 4350 GAGATCGACCTGCTGCTGAGGTGATCCAAACCCGTGGCCCTGTGATGGGAAGAGGTG 4409  
QY 361 GJuSerYrGluLeuThrValGluIleAspArgGlnGlyArgAspProGlyProArgSer 380  
DB 4410 GATCTTACCACTGACGCTGAGAGGCAAGTGAACAGGCTCGGGACCCGGGTCTCGAGAT 4469  
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
DB 4470 ACCACAGCGCTGTGTTCTCTTCTGTGAGATGACAAATGATATGCCCCCAAGTTTATG 4529  
QY 401 GJuLySArgYrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
DB 4530 GAGAAAGCTGATGTGCTCAAGGTGAGAGAGATGTGACTCCAGGGCCCCAGTACTCCGA 4589  
QY 421 ValThrIleAspAspArgAspLySArgYrSerAsnAlaValHisTyrSerIleMetSer 440  
DB 4590 GTCAAGCTCGGATCGAGACAGAGGGAGCAATGCCGTGTGCTCATATGCACTTGAATG 4649  
QY 441 GlyAsnAlaArgGlyGlnPheYrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
DB 4650 GGCATATGCTCGGGGACAGATTATCTGAGATGCCAGACTGAGACTCTGATGTGTGAGC 4709  
QY 461 ProLeuAspYrGluThrThrLySArgYrThrLeuAspValArgAlaGlnAspGlyGly 480  
DB 4710 CCTCTTGACTATGAGACGACCAAGAGTACACCTTACGGGTGCGAGCAGACAGATGTGTC 4769  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 4770 CGTCCCCCACTCTTAATGTCTGTGGCTGTGTGAGACGTACTGGAATTCACAGAC 4829  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
DB 4830 AATGCCCATCTTTCGTGACAGCCCTTTCAGGGCTACGTCTGAGAGTGTCCCTTA 4889  
QY 521 GlyYrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
DB 4890 GGTACTCTGTTCTCTCATGTCCAGGCTATCGACGCTGATGTGTGACATATCCCGCTTG 4949  
QY 541 GJuYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyYrThrGly 560  
DB 4950 GAATACCGCTTGTGTGGGTGGAGCATGACTTCCCTTCAACCATCAATGGCAGAGC 5009

QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheYrSerPheGlyVal 580  
DB 5010 TGGATCTCTGTGGCTGTGAATCGGACCGGAGAGAAATTGATTCTACGCTTTGGGGTA 5069  
QY 581 GJuIleAspAspHisGlyYrThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
DB 5070 GAGCTCAGAGACCATGGACATCCAGCACTCACTGCTGGCCAGTGAAGGTGACTGTC 5129  
QY 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluYrThrValArgLeuAsn 620  
DB 5130 CTGATGTCAACGACACCAATCCAACTTACCAACCAAGATACACAGTGGCTCAT 5189  
QY 621 GJuAspAlaAlaValGlyYrThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
DB 5190 GAGGATGACGCTGTGGGACCAAGCGTGTGACGGGTGACGCTGAGACCGTGAATGCTAT 5249  
QY 641 SerValIleThrYrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIlePheSer 660  
DB 5250 AGTGTATCACTTACCAATCAACAGTGGCAATCTCGAAACCGCTTCTCATCAACAGC 5309  
QY 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspYrYrLeuGluValArgGln 680  
DB 5310 CAAGTGTGTGGGTGTGATCTCTTCCCTTGGCCACTGAGACTCAAACTTGAAGGGCAG 5369  
QY 681 TyrValLeuAlaValThrAlaSerAspGlyYrThrArgGlnAspThrAlaGlnIleValVal 700  
DB 5370 TATGTGTGGCTGTATACCGCTCCAGTGGCACTGGCGCAGACAGGCAAGATGTGGTG 5429  
QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
DB 5430 AATGTACCGACGCGCAACCCATGCTGTCTTTCAGACTCCCACTATACAGTGAAT 5489  
QY 721 ValAsnGluAspArgProAlaGlyYrThrValValLeuIleSerAlaThrAspGluAsp 740  
DB 5490 GTTATATGAGACCGGCGGACAGCAGCAGGTGTGTGATCAAGGCCACGAGTGAAGAC 5549  
QY 741 ThrGlyGluAsnAlaArgIleThrYrPheMetGluAspSerIleProGlnPheArgIle 760  
DB 5550 ACAGGTGAAATGCGCGATCACTTCAATGAGAGACAAATGCCCATTCCTCGCATTC 5609  
QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspYrGluAspGluValSer 780  
DB 5610 GATGACAGACCGGGGCTGTACCAACCGCTGAGCTGACTATGAAGACCAATGTGCT 5669  
QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnYrSerAspThrThrYr 800  
DB 5670 TACACCTGGCCATTACTGCTCGGAGACATGCAATGCCCAAGATCGACACCACTTAC 5729  
QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgSerYr 820  
DB 5730 CTGAGATCCTGTGTGAACGAGTGAATGACATGCCCTCAGTTCCTGCGAGACTCTTAC 5789  
QY 821 GlnGlySerValYrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
DB 5790 CAGGGCAGTGTCTATGAGAGATGTGCCACCTTCACTAGCGTCTGCAATCTCAGCCACT 5849  
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheYrThrPheGlnGlyYrAspAspGly 860  
DB 5850 GATGTGATCTTGACCTTAATGAGCAGGCTTTCACACTTCAAGAGGCGAGAGATGA 5909  
QY 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880  
DB 5910 GACGGTACTTATATGTGAGTCCACGTGAGCAATCGTGCACAAAGCTACGAGGCTGAT 5969  
QY 881 ArgGluAsnValAlaGlnYrValValLeuAspArgAlaYrAlaValAspLySArgYrMetProPro 900  
DB 5970 CGAGAGAAAGTGGCCAGTATGTCTTGGCGCAATGCAAGTGAAGAGGAGATCCCCCA 6029  
QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920  
DB 6030 GCCCGCACACCTTATGAGAAAGTACAGTCACTGTGTGATGATGACAAATCCCCCTGTC 6089  
QY 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940

Db	Qy	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	6090	TTTGAGCAGAGTGGAGTTGATCTGTGTTTGTGGAAAGAAACAGCCCATTTGGGCTAGCCGTG							RESULT 10				
Qy	941	AlaArgValAlaThrAlaThrAspProAspGluGlyIleThrAsnAlaGlnIleMetCysGlnIle							LOC284647				
Db	6150	GCCCCGGGTCACAGCCACTGACCCCGCATAGAGCAACAAATGCCAATTATATGACCAAGATT							LOC284647				
Qy	961	ValGluGlyAsnIleProGluValAlaPheGlnLeuAspIlePheSerGlyIleuThrAla							DEFINITION				
Db	6210	GTGAGAGGCAACATCCCTGAGAGCTCTTTACGCTGACATCTTCTCCGGGAGACTGACAGCC							227194 bp DNA linear HTG 04-MAR-2003				
Qy	981	LeuValAspLeuAspIleProGluAspArgProGluIleValIleGlnAlaThrSer							Accession				
Db	6270	CTGGTAGACTTGAAGCTACAGAGACACCGGCTGAGTACGCTCGTGATCATCAGGCCACGTCA							Version				
Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAspPro							Keywords				
Db	6330	GCTCTCTCTGGTAGAGCCGGGCTTACAGTCCACGTCGCTCTTGACCGGCAATGACACACCA							Source				
Qy	1021	ProValIleuGlyAsnAspGluIleuPheAsnAsnIleValIleThrAsnArgSerSerSer							Organism				
Db	6390	CCAGTGTGGGCAACTTTGAGATCCTTTTCAACAACATATGTCACCAATCGCTCAAGACAGC							Reference				
Qy	1041	PheProGluGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu							Authors				
Db	6450	TTCCCTCGGGGTGTCATTGGCCGAGTACCTGCCCATGACCTGTATCTCAGATAGTCTG							Title				
Qy	1061	ThrTyrSerPheGluArgIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu							Journal				
Db	6510	ACTTACAGAGCTTTGAGCCGGGAAATGAACTACAGCTGCTGCTCAATGCTCCACAGGCT							Comment				
Qy	1081	GluLeuIleuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal											
Db	6570	GAGCTGAGAGCTTAAACCCGCGACATGGACAAACACCGGCTCTGAGAGGCATCATGAGCGTG											
Qy	1101	LeuValSer 1103											
Db	6630	CTGGTGTCTCA 6638											
Qy	941	AlaArgValAlaThrAlaThrAspProAspGluGlyIleThrAsnAlaGlnIleMetCysGlnIle											
Db	6150	GCCCCGGGTCACAGCCACTGACCCCGCATAGAGCAACAAATGCCAATTATATGACCAAGATT											
Qy	961	ValGluGlyAsnIleProGluValAlaPheGlnLeuAspIlePheSerGlyIleuThrAla											
Db	6210	GTGAGAGGCAACATCCCTGAGAGCTCTTTACGCTGACATCTTCTCCGGGAGACTGACAGCC											
Qy	981	LeuValAspLeuAspIleProGluAspArgProGluIleValIleGlnAlaThrSer											
Db	6270	CTGGTAGACTTGAAGCTACAGAGACACCGGCTGAGTACGCTCGTGATCATCAGGCCACGTCA											
Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAspPro											
Db	6330	GCTCTCTCTGGTAGAGCCGGGCTTACAGTCCACGTCGCTCTTGACCGGCAATGACACACCA											
Qy	1021	ProValIleuGlyAsnAspGluIleuPheAsnAsnIleValIleThrAsnArgSerSerSer											
Db	6390	CCAGTGTGGGCAACTTTGAGATCCTTTTCAACAACATATGTCACCAATCGCTCAAGACAGC											
Qy	1041	PheProGluGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu											
Db	6450	TTCCCTCGGGGTGTCATTGGCCGAGTACCTGCCCATGACCTGTATCTCAGATAGTCTG											
Qy	1061	ThrTyrSerPheGluArgIleGlyArgValProAlaHisAspProAspIleSerAspSerIleu											
Db	6510	ACTTACAGAGCTTTGAGCCGGGAAATGAACTACAGCTGCTGCTCAATGCTCCACAGGCT											
Qy	1081	GluLeuIleuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAl											

[illegible]



QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgLysArgAsnValAsnThrAlaProGln 180  
 DB AGGTGACGAGAGAGTCCCTGGGTGGCGCTGGAGAAAAGAAATTAAATCAGCCCCCAG 3869  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB TTCCAGCCCCCGCTGACAGGCGCACAGTCCGAGAACCCAGCCACAGGACCCCTGTT 3929  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyArgLeuGlyLysThrMet 220  
 DB GATTCCTGAGGCGCATCGACCCGAGAGGTGAGGCGAGTGACTGAGATGACACAG 3989  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 DB GATGCCCTCTTGTATAGCGCGCTCCAAACAGTTCTTCCCTGGACCAAGTCACTGGTGA 4049  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
 DB GTAAACCAAGCCGAGAGCTGATCGTGAACCAAGACCAACCCCTCTTCAAGGCTCAG 4109  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB GGCGAGGACCAACGCGCATGCCCCGACAGAGTCCCTGCTACACTCACTCATTTGGTTACT 4169  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlyLysArgLysSerLeuArgGlu 300  
 DB GACACCAATGACCATGACCTGTGTTCAGACACAGAGATCAAGAGAGCCTCAGGAG 4229  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB AACCTGAGAGGTGGCTATGAGGTCTCATCTGACAGGCCACGATGGTATATGCCCTCCC 4289  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysProSerArgValPhe 340  
 DB AATGCCAATTTCTTATCCCGCTGCTGAGAGGCTCTGGGGCAGCCCTCTGAAAGTCTTT 4349  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
 DB GAGATGACCCCTGCTCTGGGTGATCCGAAACCGTGGCCCTGTGATCGGAGAGGTG 4409  
 QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 DB GAATTCCTACAGGTGAGGAGGACAGTACCAAGGCTCGGAGCCCGGGTCTCGAGT 4469  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPheSer 400  
 DB ACCACAGCCCGCTGTTCTTCTCTGTGAGGATGACATGATTAATGCCCCCAAGTTAGT 4529  
 QY 401 GluLysArgTyrValAlaGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 DB GAGAAACGCGATGTGCTGAGTGAAGGAGAGATGATCTCAGGGGCCCACTGATCCCA 4589  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValIleHisTyrSerIleMetSer 440  
 DB GTACAGACCTCGGATGACAGCAAGGAGCAAGCGGTGTGACATATACATCAAGAGT 4649  
 QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValIleSer 460  
 DB GGCAATGCTCGGGGACAGTTTATCTGATGCCCAAGCTCGAGCTCTGATGTGGAGC 4709  
 QY 461 ProLeuAspTyrGluThrThrLysGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
 DB CCTCTTGACATATGACAGCAAGAGTACACCTTCAGGGTGCAGAGCAGATGTGGC 4769  
 QY 481 ArgProProLeuSerAsnValIleSerGlyLeuValThrValGlnValLeuAspIleAsnArg 500  
 DB CGTCCCGCATCTCTAATGCTCTGAGCTTGTGATCAAGTACAGTCTCTGATATCAAGAC 4829  
 QY 501 AsnAlaProIlePheValIleThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 DB AATGCCCCCATCTTCTGTGACACCCCTTTCAGAGCTACTGTCTGTGAGAGATGCCCTTA 4889

QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 DB GGCTACCTGGTTCTCATGTCAGGCTATCGACGTGATGCTGGTGAACAATCCCGCGCTG 4949  
 QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 DB GAAATCCCGCTTGGCTGGGTGGAGCATGATCCCTTTCACATCAACATATGACAGGC 5009  
 QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluValAspPheTyrSerPheGlyVal 580  
 DB TGGATCTCTGTGGCTGCTGATGACCGGAGAGAGTTGATTTCTTACGCTTTGGGGTA 5069  
 QY 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 DB GAAGCTCGAGACCATGGACCTCAGACCTCACTGCTCGGCGCAGTGCAGCTGACTGTC 5129  
 QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 DB CTGATGTCAACGAAACAAATCAACTTACCCAAACAGATACACAGTGGGCTCAAT 5189  
 QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 DB GAGATGCACTGTGGGACCAACGCTGTGACGAGTGTCACTGTGACCGTATGCTCAT 5249  
 QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 DB AGTGTATCACTCACTCAAGATCAACAGTGGCAATATCTCGAAACCGCTTCTCCATCAACAGC 5309  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArgGln 680  
 DB CAAGTGTGGTGGCTGGTATCCCTGCTGCGCACTGAGTACAACTTGAAGCGGCGAG 5369  
 QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 DB TATGCTGTGGCTTATACCGCTCCGATGGCACTCGGAGAGAACCGCAAGTTGGTGG 5429  
 QY 701 AsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 DB AATGTCAACGACCGCAACCACTCATGCTGTCTTTCAGAGCTCCCACTATCACTGAT 5489  
 QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 DB GTTATATGAGACCGGCGCGGACGACAGGCTGAGTGTGATAGCGCCACGATAGAGAC 5549  
 QY 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 DB AAGGTGAGAAATGCCCGATCACTTCAATGAGAGACAGCATCCCCAGTTCCGCAATC 5609  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 DB GATGACAGACCGGGGCTGTCCACCAAGCTGAGCTGAGATATGAAACCAAGTGTCT 5669  
 QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyr 800  
 DB TACACCTGCGCATTAATGCTCTGGGACATGGCATTTCCCAAGATCCGACACACTTAC 5729  
 QY 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 DB CTGAGAGATCTCGTGAACGACGTGATGACATATGCCCTCACTTCTCGGAGACTCTTAC 5789  
 QY 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 DB GAGGAGAGTGTATGAGAGATGTGCACCTTCACTGAGGTCTGTGAGATCTCAGCAT 5849  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 DB GATCGTATCTTGACATTAATGAGGAGGTCTTCTTACACTTCCAAAGAGGCGACATGTA 5909  
 QY 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 DB GACGGTGACTTATATTTGAGTCCAGTCAAGGACATCTGTGCAACCTACGAGGCTGAGT 5969  
 QY 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900



```

Db      5970 CGAAGAAACGCGCCAGTATCTTGGCGGCAATGCAAGGAGGAGGCCCA 6029
Qy      901 AAlaArgThrProMetGluValThrValThrValIleuAspValIleuAspAsnProProVal 920
Db      6030 GCCCGCACACCTATGAAAGTACAGTCACTGTTGATGATGAAATGACAAATCCCTGTC 6089
Qy      921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      6090 TTATAGCAGAGAGTGGTGGATGTTGTTGGAGAGAACACCCCATGGGCTGAGCCGTG 6149
Qy      941 AAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      6150 GCCCGGCTCACAGCACCTGACCCGATTAAGCACCAATGCCCATTTATATACCAAGTT 6209
Qy      961 ValGluGlnAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db      6210 GTGAGAGGCAACATCCCTGAGGCTTTTCAGCTGAGCATCTTCGGGAGAGCTGACAGCC 6269
Qy      981 LeuValAspLeuAspPyrGluAspArgProGluTyrValIleValIleGlnAlaThrSer 1000
Db      6270 CTGGTGAACCTTGAAGTACAGAGACCGGCTGATGACGTCCTGTCATCCAGCCACGTC 6329
Qy      1001 AAlaProLeuValSerArgAlaThrValHisValAlaProLeuAspArgAsnAspAsnPro 1020
Db      6330 GCTCTCTTGGTGAAGCCCGGCTACAGTCCAGCTCCGCTCTTGACCGCATGACACCA 6389
Qy      1021 ProValLeuGlnAspPheGluIleLeuPheAsnAsnTyrValIleAsnArgSerSerSer 1040
Db      6390 CCAAGTGTGGGCACTTGAATCATCTTTCAACAACATATGTCACATCCGTCAGAGCAGC 6449
Qy      1041 PheProGluGlyValIleGlyArgValProAlaHisAspProAspIleSerArgSerLeu 1060
Db      6450 TTCCCTGGGGGTGCATTGGCCGAGTACCTGCCATGACCTGATATCTCAGATATGCTG 6509
Qy      1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuAsnAlaSerThrGly 1080
Db      6510 ACTTACAGCTTGAAGCGGGAAATGAATCACTGAGCTGCTGCTCAATGCTCCACGGGT 6569
Qy      1081 GlnLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal 1100
Db      6570 GAGCTGAAGCTAACCCGCGCTGACACAAACCGGCTCTGAGGCGCATATAGGCTG 6629
Qy      1101 LeuValSer 1103
Db      6630 CTGGTGTCA 6638

```

```

RESULT 11
AL157901/c      116603 bp  DNA  linear  HTG 10-JUL-2001
LOCUS DEFINITION Homo sapiens chromosome 1 clone RP4-641D22 map pl3.1-13.3, ***
SEQUENCING IN PROGRESS ***, 9 unordered pieces.
AL157901
ACCESSION AL157901.5 GI:9796614
VERSION HTG: HTGS PHASE1: HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS McLay, K.
TITLE Direct Submision
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9212479.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d0641D22

```

```

----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator ABI; 3% of reads
Chemistry: Dye-terminator Big Dye; 96% of reads
Consensus quality: 112460 bases at least Q40
Consensus quality: 114130 bases at least Q30
Consensus quality: 115038 bases at least Q20
Insert size: 115803; sum-of-contigs
Insert size: 115316; 1.9% error; agarose-tp
Quality coverage: 3.82x in Q20 bases; sum-of-contigs Quality
coverage: 4.08x in Q20 bases; agarose-tp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 3349 3348: contig of 3348 bp in length
* 3349 3348: gap of 100 bp
* 3449 13632: contig of 10184 bp in length
* 13633 13732: gap of 100 bp
* 13733 20620: contig of 6888 bp in length
* 20621 20720: gap of 100 bp
* 20721 24173: contig of 3453 bp in length
* 24174 24273: gap of 100 bp
* 24274 41397: contig of 17124 bp in length
* 41398 41497: gap of 100 bp
* 41498 81263: contig of 39766 bp in length
* 81264 81363: gap of 100 bp
* 81364 92945: contig of 11582 bp in length
* 92946 93045: gap of 100 bp
* 93046 97949: contig of 4904 bp in length
* 97950 98049: gap of 100 bp
* 98050 116603: contig of 18554 bp in length.
Location/Qualifiers
source 1. 116603
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p13.1-13.3"
/clone="RP4-641D22"
/clone_1lb="RPCT-4"
1. 3348
/note="assembly_fragment:00133
fragment_chain:1"
3449. 13632
/note="assembly_fragment:01128
fragment_chain:1"
13733. 20620
/note="assembly_fragment:00187
fragment_chain:2"
20721. 24173
/note="assembly_fragment:00757
fragment_chain:2"
24274. 41397
/note="assembly_fragment:00109"
41498. 81263
/note="assembly_fragment:00393"
81364. 92945
/note="assembly_fragment:00527"
93046. 97949
/note="assembly_fragment:00530"
98050. 116603
/note="assembly_fragment:00873"
29209 a 29209 c 28975 g 28294 t 800 others

```

```

BASE COUNT 29325 a 29209 c 28975 g 28294 t 800 others
ORIGIN
Alignment Scores:

```



```

Db      67516 CAGGACACGCGACAGATTGTGGATATGTCACCGACCAACCCATGCTGCTT 67457
Qy      713 GlnSerSerHisGlyThrValAsnValAsnGluAspArgProAlaGlyThrThrValVal 732
Db      67456 CAGAGCTCCCACTATACAGTAAATGTTAAAGAGACCGGCGCGACGACCAACGCGGAG 67397
Qy      733 LeuIleSerAlaThrAspGluAspThrGlyIleAsnAlaArgIleThrTyPheMetGlu 752
Db      67396 CTGATCAGCGCCACGATGAGGACACAGGTGAGAAATGCCGATCACTCACTTCAAGAG 67337
Qy      753 AspSerIleProGluPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlu 772
Db      67336 GACAGATATCCCAAGTTCGACATCGATGACACACGCGGCGCTGTCAACCAAGCGTGA 67277
Qy      773 LeuAspArgGluAspGlnValSerTyThrLeuAlaIleThrAlaArgAspAsnGlyIle 792
Db      67276 CTGGACTACGAGACCAAGATGCTTACACCTGGCCATTAAGCTGGGACATGGCAT 67217
Qy      793 ProGlnTySerAspThrThrTyLeuGluIleLeuValAsnAspValAsnAspAsnAla 812
Db      67216 CCCCAAGATCCGACACCACTACCTGAGATCTGTGAACGATGATGACATGACATGCG 67157
Qy      813 ProGlnPheLeuAspSerTyThrGlnGlySerValTyGluAspValProProPheThr 832
Db      67156 CCTCAGTTCCTCGAGACTCTTACCAAGGCGAGTGTATGAGAGATGTGCCACCTTCACT 67097
Qy      833 SerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTy 852
Db      67096 AGCGTCTCTCAATCTCAAGCCACTGATCGTGAATCTGAGCTTAAAGGCGGCTTCTAC 67037
Qy      853 ThrPheGlnGlyIleAspAspGlyAspGlyIleValGlnSerThrSerGlyIle 872
Db      67036 ACCTTCCAAAGGAGGAGCGATGAGACGCGTGACTTATATTGTTGAGCCACGTCAGGCA 66977
Qy      873 ValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyValLeuArgAlaTy 892
Db      66976 GGGCAACGCTACGAGGCTGATGAGAGAACGTCGACAGTATGTCCTGGCGCAAT 66917
Qy      893 AlaValAspGlyGlyMetProProAlaArgThrPrometGluValThrValThrValLeu 912
Db      66916 CGAGTGGACAAAGGAGATGCCCCCAAGCCGACACCTTAAGAGTGAAGTCACTGTGTT 66857
Qy      913 AspValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGlnGlu 932
Db      66856 GATGGATGATGACAAATCCCTGCTTGAAGCAGATGATGATGATGTTGTGTGGAAG 66797
Qy      933 AsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThr 952
Db      66796 AACACCCCTATGGGTACGCGTGGCCGCGGTCAACGACCACTGACCCCATGACAGG 66737
Qy      953 AsnAlaGlnIleMetTyGlnIleValGlnGluValAsnIleProGluValPheGlnLeuAsp 972
Db      66736 AATGCCAGATTTATATACAGATTTGTGAGAGGCAACATCCCTGAGGTCTTCAAGCTGAC 66677
Qy      973 IlePheSerGlyGlnLeuThrAlaLeuValAspLeuAspTyGluAspArgProGluTy 992
Db      66676 ATCTTCTCCGGGAGACTGACACCTCGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 66617
Qy      993 ValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArg 1012
Db      66616 GTCCCTGATATCAGGCGCACGTCAGCTCTGTGGAGAGCGGCGGTCAACAGTCCAG 66557
Qy      1013 LeuLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeuPheAsnAsn 1032
Db      66556 CTCTTGAACCGAAGACAAACCAACCAAGTCTGGGCAATTTGAAATCTTTTCAACAC 66497
Qy      1033 TyrValThrAsnArgSerSerPheProGlyIleGlyValArgValProAlaHis 1052
Db      66496 TATGTACCAATCGCTCAAGACGCTTCCCTGGGCGGTGCAATTTGGCGAGTACTCCCAT 66437
Qy      1053 AspProAlaIleSerAspSerLeuThrTySerPheGluArgGlyAsnGlnLeuSerLeu 1072

```

```

Db      66436 GACCTGATATCTCAGATATGCTGACTTACAGCTTTGAGCGGGAATGAACTCAGCT 66377
Qy      1073 ValLeuLeuAsnAlaSerThrGlyIleuTyLeuSerArgAlaIleuAspAsnArg 1092
Db      66376 GTCTGTCTCAATGCTTCCACGGGTAGCTGAAGCTTAACCCGCGACGACAAACCG 66317
Qy      1093 ProLeuGlnAlaIleMetSerValLeuValSer 1103
Db      66316 CCTCTGAGGCGCATATGATGAGCTGTGCTGTC 66284

RESULT 12
AK091437
LOCUS
DEFINITION
Hom sapiens cDNA FLJ34118 f1s, clone FCBBF3009428, highly similar
to Homo sapiens FLAMINO 1 mRNA.
ACCESSION
AK091437.1 GI:21749808
VERSION
oligo capping, f1s (full insert sequence).
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T.,
Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R.,
Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, B., Momiyama, H.,
Onogawa, S., Kaeriyama, S., Satoh, N., Matsumawa, H., Takahashi, B.,
Karaoka, R., Kuge, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S.,
Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T.,
Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai, H., O., Y., Salto, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahara, K., Maehno, Y., Negai, K. and Isozaki, T.
NEO human cDNA sequencing project
Unpublished
2
(bases 1 to 2821)
Isozaki, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isozaki, FUJ Project (HRI Team) ; 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 293-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..2821
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FCBBF3009428"
/tissue_type="brain"
/clone_lib="FCBBF3"
/dev_stage="fetal"
/note="cloning vector: pMB18SFL3"
BASE COUNT
540 a 877 c 835 g 569 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2821
Score: 609.00 Matches: 782
Percent Similarity: 99.62% Conservative: 0
Best Local Similarity: 99.62% Mismatches: 2
Query Match: 20.83% Indels: 3
DB: 9 Gaps: 0
US-09-916-849a-3 (1-2923) x AK091437 (1-2821)

```

QY	1116	ArgValThrIleIleThrAspGluMetLeuThrHisSerIleIleThrLeuArgLeuGluAsp	1135
Db	221	CGTGGACCATCATCATCAACCGATGAGTGTCTCAACCCACACATCAACGCTCGCTGGAGGAC	290
QY	1136	MetSerProGluArgPheLeuSerProIleuGluGlyLeuPheIleGlnAlaValAlaAla	1155
Db	291	ATGTACCCGAGCGCTTCTGTCACTGCTAGGCTTGTTCATCCAGCGCGTGGCGCC	350
QY	1156	ThrLeuAlaThrProProAspHisValValValPheAsnValGlnArgAspThrAspAla	1175
Db	351	ACGCTGGCCACGGCACCGGACCAAGTGGTGGCTTCAACGTACAGCGGGACACCGAGCC	410
QY	1176	ProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGly	1195
Db	411	CCCGGGGGCCACATCTCAACGTGAGCTGTCTGGGGGGCCAGCGGCCCGCGGGGGC	470
QY	1196	GlyProProPheLeuProSerGluAspLeuGlnGluValGlyLeuTyrLeuAsnArgSerLeu	1215
Db	471	GGGCGGCGCTTCTGCTGCTCTGTAGGACCTGTGAGAGCGGCTTATCACTCAACCGAGCTTG	530
QY	1216	LeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsnIleCysLeuArgGlu	1235
Db	531	CTGACGGGCATCTCGGCACAGCGCGTGTGCCCTTCGACGACCAATCATGTGCTGGGGAG	590
QY	1236	ProCysGluAsnTyrMetArgCysValSerValLeuAspPheAspSerSerAlaProPhe	1255
Db	591	CCCTCGAGAACTATCATGCGCTGCGTGTGTGGGTGCTGCGCTTCGATCTCTCGCGCCCTTC	650
QY	1256	IleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArg	1275
Db	651	ATCGCTCTCTCTCCGTGCTCTTCGCGGCCATCCACCCCGTCGAGAGGCTGCGCTGCGCC	710
QY	1276	CysProProGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArg	1295
Db	711	TGCCCGCGCGGCTTACCGGGTGACTACTGCGGAGACCGAGGTGGACCTGTGCTACCTGGG	770
QY	1296	ProCysGlyProHisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArg	1315
Db	771	CCCTGTGGCCCCACAGGGGGGCTGCGCGACGCGCGGGCGGCTAACCTGTGCTGTGTGT	830
QY	1316	AspGlyTyrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGly	1335
Db	831	GATGGCTACAGGGGTGAGACCTGTAGGTGATGTCTGCTCAAGCGCTTGGAC-CCCGCT	889
QY	1336	ValCysIleAsnGlyGlyThrCysValAsnLeuLeuValGlyGlyPheArgCysAspCys	1355
Db	890	GTCTGCAGAAATGGGGGACCTGTGTCAACCTGCTGGGGGGCTTTCATGTGCATATGC	949
QY	1356	ProSerGlyAspPheGluLeuSerProTyrCysGlnValThrThrArgSerPheProAlaHis	1375
Db	950	CCATGTGAGACTTGTGAAGAGCCCTACTGCGCACAGGTGACAGCGAGCTTCCCGCCAC	1009
QY	1376	SerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPhe	1395
Db	1010	TCCCTTCATCACTTTCCGGCGCTGGGCCAGGCTTTCACCTTCACTTGGCCCTCTCGTTT	1069
QY	1396	AlaThrIleGluArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluIleHisAsp	1415
Db	1070	GCCACAAAGAGCGGACGCGGTGTCTGTGTATCAATGTGGCGTTTCAAGAAGAGATAC	1129
QY	1416	PheValAlaLeuGlnValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSer	1435
Db	1130	TTTGTGGCCCTCGAGGTATCCAGAGACGGTCCAGCTCACTTCTGTGCAAGGGAGCTCA	1189
QY	1436	ThrThrThrValSerProPheValProGlyGlyValSerAspGlyGlnIlePheHisThrVal	1455
Db	1190	ACCAACCAAGGTGTCCCATCTGTGCGCGGAGAGTCAGTGAATGGCACTGGCATACGGTGT	1249
QY	1456	GlnLeuIleTyrTyrAsnIleProIleuGluGlyIleThrGlyLeuProGlnGlyProSer	1475
Db	1250	CAGCTGAATACTCAATTAAGCACTGTGGGTGAGAGAGGGCTTCCACAGGGCCCATCA	1309

QY	1476	GIuGIuIySvAlaIaVaIaThrValAspGIyCySaaphrThrgIyValAlaLeuArgPhe	1455
Db	1310	GAGCAAGAGGCGTGTGTGACCGTGTGAATGGCTGTGAACAAGAGTGTGGCTTCGCTTC	1369
QY	1496	GIySerValLeuGIyAaenTyrSerCyaaIaIaGIuGIyThrgInGIyGIySerIyLeys	1515
Db	1370	GGATCTGTCCCTGGGCAACTACTCTGTGTCTGCCAGGGCAACCCAGGGTGGCAGCAAGAG	1429
QY	1516	SerLeuAspLeuThrgIyProLeuLeuGIyGIyValProAspLeuProGIuSerPhe	1535
Db	1430	TCTGTGTGATCGAAGGGGGCCCTCTCACTAGCGGGGGCTGTACCTGGCCGAAGCTTC	1489
QY	1536	ProValArgMetArgGIuPheValGIyCybMetArgbMetLeuGIuValAspSerArgHis	1555
Db	1490	CCAETCCGAAGACGGGCAAGTGTGTGGCTGTACATGGCAACTGTGAGTGTGAACGCCGGAC	1549
QY	1556	ILeapMetAlaAspPheILeIaAsnaaNGIyThrValProGIyCybProAlaIyLeys	1575
Db	1550	ATPAGCATGTGTGACTTCAATTGCCAACAATGGCACCGGTGCTGTGCTGCCCAAGAG	1609
QY	1576	AaenValCybAspSerAaenThrCybHisAaenGIyGIyThrCybValAaenGIuTPAspIla	1595
Db	1610	AACGTGTGTGACAGCAACACTTGCAACAATGGGGCACTTGTGTGAACAGTGGGACCG	1668
QY	1596	PheSerCySGIuCyProLeuGIyPheGIyGIySerCybaIaGIuMetAlaAaen	1615
Db	1670	TTCACTGTGCGAGTGTCCCCCTGTGGCTTTGGGGGCAAGAGCTGTGGCCCAAGAAATGGCAAT	1729
QY	1616	ProGIuNHAspLeuGIySerSerLeuValAlaATPHisGIyLeuSerLeuProIleSer	1635
Db	1730	CCACAGCACTTCCGTGGGAGAGAGCGCTGTGTGGCCCTGTGACATGTGCTCC	1789
QY	1636	GIuProTPryIyLeuSerLeuMetPheArgThrArgInaIaAspGIyValLeuLeuGIuIn	1655
Db	1790	CAACCTGTGTACTCAAGCTCTATTTCCGCAACGGGCCAGGGCCAGCGGTCTCTGTGAG	1849
QY	1656	AlaIleThrArgGIyArgSerThrIleThrLeuGIuLeuArgGIuIyHisValMetLeu	1675
Db	1850	GCCATTCACAGAGGGGGCGCAGACCATCATCACTTACAGCTACAGAGGGGCGACGTATGTGTG	1909
QY	1676	SerValGIuGIyThrgIyLeuGIuAlaSerSerLeuArgLeuGIuProGIyArgAlaAaen	1695
Db	1910	AGCGTGAAGGCAACAGGGCTTCAGAGCTCTCTCTCTCGGTCTGTGAAGCAAGCGGGCCAAAT	1968
QY	1696	AspGIyAspTPHisHisIaIaGIuLeuAlaLeuGIyAlaSerGIyGIyProGIyHisIaI	1715
Db	1970	GACGGTAGCTGTGCACATGACG-GCTGTGCACTGTGGAGCGACGGGGGGCCGGCCATGTC	2028
QY	1715	AlaIleLeuSerPheAspTyrgIyGIuGIuInaArgAlaGIuGIyAaenLeuGIyProArgLeuHis	1735
Db	2039	CATTCTGTCTCTTCATTATATGGGCAAGAGAGAGAGAGGCAACTGTGGCCCCCGAGCTGTGA	2088
QY	1735	sgIyLeuHisIeLeuSerAaenIleThrValGIyGIyIleProGIyProAlaGIyGIyValAl	1755
Db	2089	TGTGTGTGACCTGAGCAACATACATACATGTGGCGGAATCTGTGGCCAGCGCGGGTGTGTGC	2148
QY	1755	AaArgGIyPheArgGIyCybLeuGIuGIyValArgValSerAspThrProGIuGIyValAs	1775
Db	2149	CCGTGTGCTTTCGGGGGCTGTGTGGCAAGGTGTGTGGGTGTGAAGCATACGCCCGAGGGGGGTAA	2208
QY	1775	nSerLeuAspProSerHisGIyGIuGIuSerIleAaenValGIuGIuGIyCybSerLeuProAs	1795
Db	2209	CAGCTGTGATCCAGCCATGTGGAGAGAGCATACGTGAGAGCAAGCTGTAGCTGTGCTGA	2268
QY	1795	pProCybAspSerAaenProCybProAlaAaenSerTyTCySerAaenAspTPAspSerTy	1815
Db	2269	CCCTTGTGTGACCAAAACCGGTGTCTGTGTAAACAGCTATGTGCAACACATGGGACAGCTA	2328
QY	1815	rSerCybSerCybAspProGIyTyThrTyrgIyAaenAspCybThzAaenValCybAspLeuAs	1835
Db	2339	TTCTGTGAGCGTGTATCCAGGTATCTATGTGTGAACACTGTATGTATGTGTGTGACCTGTGA	2388
QY	1835	nProCySGIuHisGIuSerValCybThzArgIyPProSerAlaProHisGIyTyThrTyC	1855

Db 2389 CCGGTGTGACACCACTGTGTGTACCCGCAAGCCAGTCCCAAGGCTATACCTG 2448  
 QY 1855 GGLUCYSPROPTAMTYRLEUGLYPROLYTYRGSLUTRARGILEASPLNPROCYSPR 1875  
 Db 2449 CCGAGTCTCCCAATTAATCTTGGCCATACCTGTGAGACCAAGATTGACCACTTGTCC 2508  
 QY 1875 CARGLYTPTTPTGLVHISPROTHRCYSGLYPROCYSAANCYASPAVALSERLYSGLYPh 1895  
 Db 2509 CCGTGTCTGTGTGGGACATCCCAATGTGCCATGCACTGATGTGACGAAAGCTT 2568  
 QY 1895 EASPPROAPPOCY 1899  
 Db 2569 TACCCAGACTGC 2581

RESULT 13  
 AL672200/c  
 LOCUS AL672200 89129 bp DNA linear ROD 01-JUL-2002  
 DEFINITION Mouse DNA sequence from clone RP23-37G19 on chromosome 3, complete  
 sequence.  
 ACCESSION AL672200  
 VERSION AL672200.10 GI:21684704  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 89129)  
 GARNER, P.  
 DIRECT SUBMISSION  
 Submitted (01-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Jul 2, 2002 this sequence version replaced GI:21304370.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-37G19 is  
 from the R0C1-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6

COMMENT  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk

FEATURES  
 source  
 ----- Location/Qualifiers  
 1. 89129  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="3"  
 /clone="RP23-37G19"  
 /clone\_id="RP23-37G19"  
 /clone\_id="RP23-37G19"  
 2218 a 22283 c 22360 g 21268 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0  
 Score: 482.00  
 Percent Similarity: 98.86%  
 Best Local Similarity: 98.86%  
 Query Match: 16.49%  
 DB: 10  
 Gaps: 0

US-09-916-849A-3 (1-2923) x AL672200 (1-89129)

QY 385 ValPheLeuSerValGluAspAspAspAspAlaProGlnPheSerGluValArgTyr 404  
 Db 21678 GTTTCCTGTGTGGTGGAGATGATATATGACACGCCCCCAATTATGACAGAGGGTTAT 21619  
 QY 405 ValValGlnValArgGluAspValThrProGlyAlaProValLeuArgValThrAlaSer 424  
 Db 21618 GTGTCTCAGTGTGGGAGATGTGACCCAGAGGCCCAAGTTCTCCAGTCAACGGCTCC 21559  
 QY 425 AspArgAspLeuGlySerAspAlaVal-ValHisTyrSerIleMetSerGlyAspAlaAr 444  
 Db 21558 GATAGAGCAAGGCGCAGCAATGCCCT-AGTGCATACAGCATCATGAGTGCATATGCTCG 21500  
 QY 444 GGLYGLINPHERYRLEUASPAALAGINTHGLYALALEUASPAVALISERPROLEUASPT 464  
 Db 21499 GGGCCAGTTCTATCTGATGCTCAGACCTGAGCCCTGATGTGTAAAGCCCACTGACTA 21440  
 QY 464 rGluThrThrLysGluTyrThrLeuArgVal-ArgAlaGlnAspGlyGlyArgProProL 484  
 Db 21439 TGAAGCAACCAAGATATACCTTCGGAT-CCGGGCTCAGGACCGTGGCTCTCCAC 21381  
 QY 484 euserAspValSerGlyLeuValThrValGlnValLeuAspIleAspAspAlaProI 504  
 Db 21380 TCTCCAAATGCTCTGCGCTAGTAAAGCAAGTCTTACATCAACATTAACGCCCCA 21321  
 QY 504 IepValSerThrProPheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuV 524  
 Db 21320 TCTTGTGACACCCCTTTCACAGCCGCTCTTGAAGAGTGTCTCTTAAAGCTTACCTTG 21261  
 QY 524 alLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeuGluTyrArg- 543  
 Db 21260 TTCTCCAGTCCAGGCAATTGATGCTGACGCTGGTGAACGCCCGCTCCAGATTAAG-C 21202  
 QY 544 LeuAlaGlyValGlyHisAspPheProPheThrIleAspAspGlyTyrGlyTyrIleSer 563  
 Db 21201 CTGGCAGGGGTGGGACGACATTTCTCCCTTCACTTAACATTAAGGACAGGCTGATCTCT 21142  
 QY 564 ValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyValGluAlaArg 583  
 Db 21141 GTGGCCGAGAGTTGGATCGGAAGAGGTGATTTCTACAGCTTGTGTAGAGGCCCGG 21082  
 QY 584 AspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal-LeuAspVa 603  
 Db 21081 GACCATGACACCCCAAGCCCTCAGGCCCTCAGCAGTGTGATGATACAT-CTTGAGCT 21023  
 QY 603 IAspAspAspAspProThrPheThrGlnProGluTyrThrValArgLeuAspAla 623  
 Db 21022 CATATGACAAACCAACCACTTCAACGACAGAGTACACCGTTGGCTCAACGAGAGCGC 20963  
 QY 623 AlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValI 643  
 Db 20962 GCGCCGTGGACCAAGTGTGTGACCGTGTACGCCGTGATGACAGCGCTCAACAGTGTCT 20903  
 QY 643 eThrTyrGlnIleThrSerGlyAspThrArgAspArgPheSerIleThrSerGlnSerG 663  
 Db 20902 TACCTACCAAGATCACCAAGCGGACACCGGACACCGCTTCTTATCAACAGCAAGTGG 20843  
 QY 663 YGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluArgGlnTyrValLe 683  
 Db 20842 TGGTGAAGTGTCTCTCTTGTGCTTACCGCTGACATCAAACTCCAGCGCGCATGATGTCT 20783  
 QY 683 uAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValAlaAspValTh 703  
 Db 20782 GGCACTGACTGCTCATGATGGACCAAGGACGACACAGCTCATGATGATGATGATCTAC 20723

QY 703 rAapAlaantHrHlaArgProValPheGInserSerHlaTyThrValAsnValaengl 723  
 Db 20722 CGATGCCAACACCCACGCTCCGCTTCCAGAGCTCCCACTACACAGTCATATATATA 20663  
 QY 723 uAapAaGProAlaGlyThrThrValValleuIleSerAlaThrAspGluAapThrGly 743  
 Db 20662 AGACCCGCGCAGGAGGACCAAGTGTGTGATCAGCGCCACGAGTGAAGACCGGGGA 20603  
 QY 743 uAanAlaArgIleThrTyRPhMetGluAapSerIleProGlnPheArgIleAspAlaAs 763  
 Db 20602 GAATGCCCAATCACTACTTATGAGAGATACATCCCACTTCCGAATCAACCGGA 20543  
 QY 763 pThrGlyAlaValThrThrGlnAlaGluLeuAapTyrgIuAapGlnValSerTyThrIle 783  
 Db 20542 CACTGGGGGCTGTACACACCCAGGCTGAGCTACAGAGACCAAGTGTCTTATACCT 20483  
 QY 783 uAlaIleThrAlaAArgAspAnGlyIleProGlnIuAapThrThrTyrluGluIuI 803  
 Db 20482 GGCATCACTGTGCGGACATGGCAATGCCCAAGAGTGCACACCACTATTGGAGAT 20423  
 QY 803 eLeuValaAsnAapValaAsnAapAnAlaProGlnPheLeuAArgAspSerTyrgInGlySe 823  
 Db 20422 CCGTGGATGAGCTGAACACACCCCGCAGTCTCAGAGATTCCTACACAGGGCAG 20363  
 QY 823 rValTyrgIuAapValProProPheThrSerValleuGlnIleSerAlaThrAspArgAs 843  
 Db 20362 CGCTACGAGAGAGTCCCGCCCTTCAACAGCGTCTGACAGATCTCAGCCACGAGTCA 20303  
 QY 843 pSerGlyLeuAanGlyAArgValAetherTyThrPheGlnGlyValAapAspGlyAapGlyAs 863  
 Db 20302 CTCGGTCTGAAGAGGAGGGGTTTTCTACACCTTCAAGAGAGATGATGGGACGCTGA 20243  
 QY 863 pPheIleValGluSerThrSerGlyIleValAArgThrLeuAArgIleuAapArgGluAs 883  
 Db 20242 CTTTATGTAGATCGACATCGGCGCATGTGGCCACATCGCGAGCTGATGTAGAA 20183  
 QY 883 nValAlaGlnTyrValleuAArgIaTyraValaAapIyGlyMetProProAlaArgTh 903  
 Db 20182 CGTGGCCAGTACGTCTGCGGCGCTATGCGGTGACAGAGGCGCATGCCACCGCGCAC 20123  
 QY 903 rPProMetGluValThrValThrValleuAapValaAsnAapProProValPheGluIuI 923  
 Db 20122 ACCCAAGAGAGTACAGTATCTGTCTCGAGTGTGAAGTGAACACCCCGTGTCTTAACA 20063  
 QY 923 nAapGluPheAapValaPheValGluGluAanSerProIleGlyLeuAlaValaArgVa 943  
 Db 20062 GATGAGCTTGAAGTATTTCTGGAAGAGACACCCCATCGGCTGGCCCGCGGT 20003  
 QY 943 lThrAlaThrAapProAapGluGlyThrAanAlaGlnIleMetTyrgInIleValGluI 963  
 Db 20002 CACAGCCACAGCCACGATGAGAGGACCAACGACAGATCATGTACAGATGTGAGAGG 19943  
 QY 963 yAanIleProGluValPheGlnIleuAapIlePheSerGlyGluLeuThrAlaIleuVala 983  
 Db 19942 CAATATCCCTGAGAGTCTTCAAGTGAATATCTCTGCGAGCTACATGCCCTGGTGA 19883  
 QY 983 pLeuAapTyrgIuAapArgProGluTyrValleuValIleGlnAlaThrSerAlaProle 1003  
 Db 19882 TTTGAGACTAGAGGACCGGCTGAATATGTCTGTGTATCCAGGCTACGCTCTCCTT 19823  
 QY 1003 uValSerAArgAlaThrValHlaValaArgLeuLeuAapArgAsnAapAnProProValle 1023  
 Db 19822 GGTGAGAGAGGCTACGTCTCAATGCCCTCTCTTGACCGGACGATATACCCCGCGGT 19763  
 QY 1023 uGlyAanPheGluIleLeuPheAanAenyTyraValThrAanArgSerSerPheProG 1043  
 Db 19762 GGGCACTTTGAGATTTCTCTTAACACTATGTACACACCGCTCCAGACCTTCCCGG 19703  
 QY 1043 yGlyAlaIleGlyAArgValProAlaHlaAapProAapIleSerAapSerLeuThrTyse 1063  
 Db 19702 TGGTGCCATAGGCGGTGTGCTGCCACGACCCGATATCTCAGACAGCTCAGTACAG 19643

QY 1063 rPheGluAArgIyAanGluLeuSerLeuValleuLeuAanAlaSerThrGlyGluLeu 1082  
 Db 19642 CTCGAGCGAGGAATGAATCACTCACTGTGCTCAATGCTCCACCGGTGAGCTG 19585  
 RESULT 14  
 AC093365/C  
 LOCUS  
 DEFINITION  
 AC093365  
 Mus musculus clone RP23-60016, WORKING DRAFT SEQUENCE, 7 unordered  
 pieces.  
 AC093365  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 234932)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-60016  
 Unpublished  
 2 (bases 1 to 234932)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Bana, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamet, A., Karatas, A., Kelle, C., Lacroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McGowan, P., McKernan, K., McPheeler, R., Meldrum, J.,  
 Menus, L., Miroya, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Nobu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Riebeck, M., Riley, R., Rie, C., Rogov, P.,  
 Roman, J., Roselt, M., Roy, A., Santos, R., Schauer, S., Schnpack, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submision  
 Submitted (20-AUG-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 234932)  
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Bana, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choquel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Fairo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamet, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
 Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Miroya, T.,  
 Mienga, V., Miya, T., Naylor, J., Nguyen, C., Nicol, R., Nobu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunhang, P., Pierre, N., Raymond, C., Retta, R.,  
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnpack, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K.,  
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submision  
 Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 21, 2002 this sequence version RepeatMasker:  
 All repeats were identified using RepeatMasker:  
 Smit, A. F. A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center

TITLE  
 JOURNAL  
 COMMENT

QY	703	138953	138894
Db	139013	GGCAGTACATGCTCCAGATGGCACAAGGCGAGGACAGCTCAGATGATGAGTGAATGCAC	138954
QY	703	TAAPALASANTHrHisArgProValPheGlnSerSerHisrYrThrValaIaenIaIaenI	723
Db	138953	CGATGCCAACACCCACGATCCCGCTTCCAGAGCTCCCACTACACAGTCAACGTTAATGA	138894



QY	723	AApApRprroalagIyThrThValValIleuIleSerAlaThrAspGluuAspThrGluYgl	743
Db	138893	AGACCGGCGACGAGGCGACCAAGTGGTGTGATCGAGCGCACGGAAGAGGACCGGGGGA	138834
QY	743	uAsnAlaArgIleThrTyR.PheMetGluAspSerIleProGlnPheArgIleAspAlaAs	763
Db	138833	GAATGCCGAATCACCTACTTATATGGAGATAGCATCCCCAGTTCCGATGACCGGGA	138774
QY	763	pThrGluValaValIThrThrGlnAlaGluIleuAspTyrgIuAspGlnValSerTyRThIle	783
Db	138773	CACGGGGCGTGCACACCCAGAGGTGAGTGCATACGAGAGACAGATGTCTTATACCTT	138714
QY	783	uAlaIleThrAlaAcgAspAsnGlyIleProGlnTySAspThcThTyRleuGluuI	803
Db	138713	GGCCATCACTGTCGGGACAAATGGCATTTCCCGAAGCTGTGACACCACTATTGGAGAT	138654
QY	803	eIeuValAsnAspValAsnAspAsnAlaProGlnPheIeuArgAspSerTyrgInglySe	823
Db	138653	CCTGTGTGATGACGTGAACGAGAACCCGCCCGCTCTCAGAGATTCTTACACAGGGCGAG	138594
QY	823	rValTyrgIuAspValProProPheThrSerValIleuGlnIleSerAlaThrAspArgAs	843
Db	138553	CGTCTACGAGGACGCGCCGCTTACACGAGCTCTGCAGATTCAGACCGAGATGAGA	138534
QY	843	pSerGlyIleuAsnGlyArgValPheTyRThrPheGlnGlyGlyAspAspGlyAspGlyAs	863
Db	138533	CTCCGGTCTGAATGGAGGGGTTTTTCTACACCTTCCAAAGGAGAGATGATGGGAGCGGTGA	138474
QY	863	pPheIleValGluSerThrSerGlyIleValArgThrIleuArgTyrgIeAspArgGluAs	883
Db	138473	CTTATATGTAGGTGACATCGGGCAATTTGTGGCACACTCGCGAGGCTGTATGTAGAA	138414
QY	883	nValAlaGlnTyRValIleuArgAlaTyRAlaValaIAspTyrgIuMetProProAlaArgTh	903
Db	138413	CGTGCCCGCAGTACGTCCTCGGGCCATATGCGGTGACAAAGGCGATGCGACACGCCGAC	138354
QY	903	rProMetGluValIThrValIThrValIleuAspValaAsnAspAsnProProValPheGluG	923
Db	138353	ACCCTAGGAAGGACAGTACTGCTCGATGTGAATGACAAACCCCGCTGCTTGAACA	138294
QY	923	nAspGluPheAspValPheValGluGluuAsnSerProIleGlyIleuAlaValAlaArgVa	943
Db	138293	GGATAGATTGACGATTTTGTGGAGAGAAACAGCCCCCATGCGGCTGGCGGCGCGGCT	138234
QY	943	IThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrgInIleValGluG	963
Db	138233	CACAGCCACTGACCCGATGAAGAGGACCAACGACAGATCATGTACAGATTGTGGAGGG	138174
QY	963	yAsnIleProGluValPheGlnIleuAspIlePheSerGlyGluIleThrAlaIeuValAs	983
Db	138173	CAATATCCCTGAGGCTTTTCAGTTGGATATATCTCTGGGAGACTTACATGCCCTGTGAAG	138114
QY	983	pIeuAspTyrgIuAspArgProGluTyRValIleuValIleGlnAlaThrSerAlaProIe	1003
Db	138113	TTTGGACATRAGAGACCGGCGCTGAATATGTCCGTGTCAATCCAGGCTACGTCTCTCTT	138054
QY	1003	uValSerArgAlaThrValaIleValaIArgIleuIleuAspArgAsnAspAsnProProValIe	1023
Db	138053	GGTGGACAGGGCTACTGTCCATGTCCGCTCTCTTGACCGGAAAGTAACCGGCCGGGCT	137994
QY	1023	uGlyAsnPheGluIleIeuPheAsnAsnTyRValIThrAsnArgSerSerSerPheProG	1043
Db	137993	GGGCAACTTGAAGATTCTCTTAAACAATGTGTACCAACCGCTCAAGACGCTTCCCCGG	137934
QY	1043	yGlyValaIleGlyArgValProAlaHisAspProAspIleSerAspSerIeuThTyRSe	1063
Db	137933	TGGTGCATAGGCGCGTGTGCTGCCACGACCGCCGATATCTCAGACAGCGCTGACGTACAG	137874
QY	1063	rPheGluArgGlyAsnGluIleuSerIeuValIleuIleuAsnAlaSerThrGluYglIleu	1082
Db	137873	CTTCAGCGAGAAATGAACCTCAGCTTGTCTGTCTCAATGCTTCCACCGGTGAAGCTG	137816
RESULT 15			

AX098223	LOCUS	AX098223	3912 bp	DNA	linear	PAT 30-MAR-2001
DEFINITION	Sequence 135 from Patent WO0118542.					
ACCESSION	AX098223					
VERSION	AX098223.1	GI:13515345				
KEYWORDS						
SOURCE						
ORGANISM	Homo sapiens (human)					
REFERENCE	Homo sapiens					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
	Lee,J., Thompson,P. and Lillie,J.					
JOURNAL	Identification, assessment, prevention, and therapy of ovarian					
FEATURES	Patent: WO 0118542-A 135 15-MAR-2001;					
source	Millennium Predictive Medicine, Inc. (US)					
	Location/Qualifiers					
	1..3912					
	/organism="Homo sapiens"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:9606"					
BASE COUNT	763 a 1256 c 1032 g 860 t					1 others
ORIGIN						
Alignment Scores:						
Pred. No.:	0	Length:	3912			
Score:	347.00	Matches:	376			
Percent Similarity:	99.47%	Conservative:	1			
Best Local Similarity:	99.47%	Mismatches:	0			
Query Match:	11.87%	Indels:	2			
	6	Gaps:	0			
US-09-916-849A-3 (1-2923) x AX098223 (1-3912)						
QY	2186	ArgGyTgtGuaAlaLeuAagGlyGluGlnProProCaaAspLeuGluGlnThrThrValIleLeuPro	2205			
Db	11	CGTACGAGAGGCGCTGCGTGGAGACACGCCCGGACCTTGAGACAAACGATCTCGCT	70			
QY	2206	GluSerValPheArgGluThrProProValValArgProAlaGlyProGlyGluAlaGln	2255			
Db	71	GAGCTGTCTTTCAAGAGAGAGCGCCCCCGGTGTCAAGCCCGAGGCCCGCGAGAGAGGCCAG	130			
QY	2226	GluProGluGluLeuAlaAgaGlnArgArgHisProGluLeuSerGlnGlyGluAla	2245			
Db	131	GAGCCAGAGAGCGTGGACCGGCGACAGCGCACCCGAGCTGAGCCAGAGGAGGCT	190			
QY	2246	ValAlaSerValIleIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspPro	2265			
Db	191	GTGGCCAGCGTCATCATCTACCGCACCTGGCCGGGCTACGCTCATTAATGACCCCT	250			
QY	2266	AspIlyValArgSerLeuArgValProIlyArgProIleIleAsnThrProValValSerIle	2285			
Db	251	GACAGAGCGCAAGCTTGAAGTCCCAAAAGCCCGCATCTCAACACACCCGTGTAGCATC	310			
QY	2286	SerValHisAspArgGluGluLeuLeuProArgAlaLeuAspIlySerProValThrValGln	2305			
Db	311	AGCGTCCATGTATGTATGAGAGCTTCTGCCCGGCGCTTGACAAACCGTACAGGTGAG	370			
QY	2306	PheArgLeuLeuGluThrGluGluArgThrIlyAspProIleCyValPheTrpAsnHisSer	2355			
Db	371	TTCCGCCCTGCTGGAGACAGAGAGCGGACCAAGCCCATCTGTGCTCTTGAAACCATTC	430			
QY	2336	IleIleuValSerGlyThrGlyGlyTyrTrpSerAlaArgGlyCyArgIlyValValPheArgAsn	2345			
Db	431	ATCTGTGTGACATGGCACAGGTGTGTGTGGCCAGAGAGGTGTGAAGTGTCTTCCGCAT	490			
QY	2346	GluSerHisValSerCyGlnCyAsnHisMetThrSerPheAlaValIleuMetAspVal	2365			
Db	491	GAGAGCCACGTCAAGCTGCCAGTGCACCACTACGAGCTTCGTGTGTCTCATGGAGCTT	550			
QY	2366	SerArgArgGluAsnGlyGlyIleLeuProLeuIlyThrLeuThrTyrValAlaLeuGly	2385			
Db	551	TCTGGGGCGGAGATGGGAGATCTCTGCATGTAAAGCATGAGCATATCGTGGCTCTTAGGT	610			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 05:37:06 ; Search time 1422 Seconds

(without alignments)  
5548.845 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923

Sequence: 1 MRSPTGVPLPPPPPLLL.....ACTVDEDSGSRFLFPFLH 2923

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO-US09916849/runat.11022004.143916.21346/app.query.fasta.1.3079  
-DB=N Geneseq.19Jun03 -QFMT=fastap -SUFFIX=Oligo.rng -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849 @CGN 1 1 0 @runat.11022004.143916.21346 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq.19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	2923	100.0	8772	22	AA511678	Human Flamingo cDN
2	2923	100.0	10531	25	AB242868	Human GPCR CELSR2
3	2894	99.0	11762	25	ABX34546	Human mdtl cDNA SE
4	2837	97.1	8871	22	AA511677	Human Flamingo cDN
5	1977	67.6	9401	22	AB080648	Human FLAMINGO 1 h
6	1958	67.0	9321	22	ABK15177	Human REPT9 cDNA
7	1623	55.5	9121	21	AACT6401	Human ORFX ORF1956
8	776	26.5	2332	22	AAK83060	Human immune/haema
9	776	26.5	2332	22	AAK83060	Human immune/haema
10	776	26.5	2332	22	AA531490	Human DNA for a no
11	776	26.5	2332	24	AB066814	Human polynucleoti
12	729	24.9	2391	25	ABT31943	Human breast cance
13	691	23.6	2077	22	ABAI9448	Human nervous syst
14	691	23.6	2077	22	AAK83061	Human immune/haema
15	691	23.6	2077	22	AA531491	Human DNA for a no
16	691	23.6	2077	24	AB066815	Human polynucleoti
17	537	18.4	2603	19	AAV07219	Human calcitonin r
18	409	14.0	1734	22	ABAI9445	Human nervous syst
19	347	11.9	3912	22	AAK83062	Human late stage o
20	281	9.6	2695	24	ABF98728	Human polynucleoti
21	281	9.6	4152	21	AAFI5924	Human prostate can
22	220	7.5	3620	25	ABZ6319	Human secretory po
23	217	7.4	652	22	ABAI9446	Human nervous syst
24	217	7.4	652	22	AAK83062	Human immune/haema
25	217	7.4	652	22	AA531492	Human DNA for a no
26	217	7.4	652	24	AB066816	Human polynucleoti
27	177	6.1	551	21	AAAI4996	Human secreted exp
28	131	4.5	509	22	ABAI9449	Human nervous syst
29	131	4.5	509	22	AAK3055	Human immune/haema
30	131	4.5	509	22	AA531489	Human DNA for a no
31	131	4.5	509	24	AB066813	Human polynucleoti
32	118	4.0	658	22	AAI99622	Human expressed po
33	118	4.0	658	22	AA531252	Human cDNA encodin
34	118	4.0	658	24	AB066576	Human polynucleoti
35	105	3.6	443	22	ABAI3396	Human nervous syst
36	105	3.6	443	22	AA531400	Human cDNA encodin
37	105	3.6	443	24	AB066724	Human polynucleoti
38	102	3.5	721	22	ABAI9448	Human genomic DNA
39	102	3.5	721	22	ABAI9448	Human genomic DNA
40	102	3.5	721	22	AA529019	Genomic sequence #
41	102	3.5	721	22	AA529019	Genomic sequence #
42	102	3.5	721	22	AA531656	Genomic sequence #
43	102	3.5	721	22	AA531657	Genomic sequence #
44	102	3.5	721	24	ABV84182	Human polynucleoti
45	102	3.5	721	24	ABV84183	Human polynucleoti

#### ALIGNMENTS

RESULT 1  
AA511678 standard; cDNA; 8772 BP.

AA511678;

24-OCT-2001 (first entry)

Human Flamingo cDNA splice variant.

Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
obesity; hypotension; hypertension; urinary retention; angina pectoris;  
myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
benign prostatic hypertrophy; psychotic disorder; neurological disorder;  
manic depression; delirium; dementia; severe mental retardation; ss;

KM Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KW antifungal; antiviral; antiprotzoal; anti-HIV; anorectic; antianorectic;  
 KW antiparkinsonian; cardiatic; cerebroprotective; neuroprotective;  
 XX antidepressant; anticonvulsant; antisense therapy; gene therapy.

OS Homo sapiens.

PH Key Location/Qualifiers  
 FT CDS 1..8772  
 FT /tag= a  
 FT /product= "Human Flamingo protein #2"

XX MO200161003-A1.

XX 23-AUG-2001.

XX 19-FEB-2001; 2001MO-GB00680.

XX 19-FEB-2000; 2000GB-0004196.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Testa TT;

XX WPI: 2001-502792/55.

XX P-PSDB; AA007054.

XX An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -

XX Claim 5; Page 29-33; 66pp; English.

XX The sequence represents a cDNA splice variant which encodes a human  
 CC Flamingo polypeptide. Flamingo is a member of the G-protein coupled  
 CC receptor family, which is involved in signal transduction pathways. By  
 CC screening to identify compounds that stimulate or inhibit the function or  
 CC level of the protein, treatments can be developed for various diseases  
 CC and bacterial, fungal, protozoan and viral infections, including HIV,  
 CC cancer, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's  
 CC disease, acute heart failure, hypotension, hypertension, urinary  
 CC retention, angina pectoris, myocardial infarction, stroke, ulcers,  
 CC allergies and benign prostatic hypertrophy. Also treatable are psychotic  
 CC and neurological disorders such as anxiety, schizophrenia, manic  
 CC depression, delirium, dementia, severe mental retardation, Huntington's  
 CC disease and Gilles de la Tourette's syndrome.

XX Sequence 8772 BP; 1696 A; 2814 C; 2564 G; 1698 T; 0 other;

XX Alignment Scores:

Pred. No.: 0 Length: 8772  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservativity: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-916-849a-3 (1-2923) x AAS1678 (1-8772)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 DB 1 ATGGGAGGCCCGGACCGGCGTCCCTCCCAAGCCGCGCGCGCTGCTGCTG 60  
 QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyValAspGlyValGlyProCySarSerLeu 40  
 DB 61 TTGCTGCTGCTGCTGCGCGCGCACTATTGGGAGACCAAGTGGGGCCCTGCTGCTT 120  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAlaCysAlaProMetGlyTyrPleuCySarSer 60  
 DB 121 GGGGCCAGGGGACGAGCTCTTGGGGGCGCTGCGCCCAATGGGCTGCTGCTGCAATCC 180  
 QY 61 SerAlaSerLeuLeuTyrPleuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80

DB 181 TAGGCTGGAACCTTGCGCTTACACAGCCGCTGCAAGGATGCGGCACTGACTACT 240  
 QY 81 GlyHsLeuValProHsHsAspGlyLeuArgValTyrCySarProGluSerGluAlaHs 100  
 DB 241 GGCACCTGTGATCCCAACAGATGGCGTGAAGGTTTGGTGTCCAGAAATCCAGGCCCAT 300  
 QY 101 IleProLeuProProAlaProGluGlyCySarProTyrSerCySarGluLeuGlyIleGly 120  
 DB 301 ATTCCCTTACCAACAGCTTCTGAAAGGCTGCCCCCTGAGCTGTGCTCTGAGCATTTGA 360  
 QY 121 GlyHsLeuSerProGluGlyValSerLeuThrLeuProGluGluHsIleProCySarLeuVal 140  
 DB 361 GGCACCTTTTCCCAACAGGCAAGCTTCACTGCCCCGAGAGACACCTGCTTAAAGGCT 420  
 QY 141 ProArgLeuArgCySarGlnSerCySarValSerValGlnAlaProGlyLeuArgAlaGlyIle 160  
 DB 421 CCACGGCTCAGATGCAAGCTCTGCAAGCTGGGACAGGCCCGCGGCTCAGGGGAGGGA 480  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgValArgAspValAsnThrAlaProGln 180  
 DB 481 AGGTCAACCAAGAAATCTCTGGGTGGGTGGGAAAGAAATGTAATACAGCCCCCAG 540  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 541 TTCAGCCCCCAAGCTTACAGGCAAGTCCGCGAGAACACAGCCAGCAGGACCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyValArgLeuGluTyrThrMet 220  
 DB 601 GCATCCCTTGAAGGCATCGAACCCGAGAGGGTGAAGGAGCTGCACTGGAGTACACCATG 660  
 QY 221 AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyIle 240  
 DB 661 GATGCCCTTTGATAGCGCTCAACCAAGTCTTCTCCAGACCAAGTCACTGAGTGA 720  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluTyrThrLysSerThrHsIleValPheArgValThr 260  
 DB 721 GTACCAACAGCCGAGAGAGTGTGATGTGAGACCAAGACCAAGCAAGTCTTCAAGGTCACG 780  
 QY 261 AlaGlnAspHsGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 781 GCGCAGAGACCAAGGATGCCCCGAGCAAGTGCCTGTGCTACACTCACTCTTGTGTACT 840  
 QY 281 AspThrAspAspHsAspProValPheGluGlnGlnGluTyrLysGlnSerLeuArgGlu 300  
 DB 841 GACACCAATGACCAATGACCTGTGTTCAGACAGAGATACAGAGAGAGCTTCAGGGAG 900  
 QY 301 AsnLeuGluValGlyTyrGluValIleuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB 901 AACCTGAGAGTTGGCTATAGAGTGTCTCACTGTCAAGGCCAAGATGTATGCCCTGCC 960  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysProSerGluValPhe 340  
 DB 961 AATGCCAATATCTGTACCGCTGCTGAGGGGCTCGGGGGGACGCCCTCTAAGCTTTT 1020  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 DB 1021 GAGATCGACCTCGCTGCTGGGGGTGATCCCAACCCGCGCTGTGTATGGGAAAGGGTG 1080  
 QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 DB 1081 GATCTTACACACTGACGCTTAAAGCAAGTGAACAGGCTCGGAGCCCGGCTCTGAGAT 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 DB 1141 ACCACAGCCCGCTGTTTCTTCTCTGTGGAGATGCAATGAATAAATGCCCCCAAGTTTACT 1200  
 QY 401 GluLysArgTyrValValGlnValArgGluAspValThrProGlyValaProValLeuArg 420  
 DB 1201 GAGAGCGCTTATGTGTGTCAGGTGAGGAGATGTGATCTCCAGGGGCCCAATCTCCA 1260  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValIleGlySerIleMetSer 440  
 DB 1261 GTCAAGCTCTCGATCGATGAGACAAAGGAGCAATATGCCGTGTGTCACTATGCAATCATGAGT 1320

QY	441	GLYAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaLeuAspValAlaSer	460
Db	1321	GGCAAGCTCGGGGAGAGATTATCTGGATGCGCCAGACTGAGCTCTGATGTGGAGAC	1380
QY	461	ProLeuAspDyGlnThrTlySGlyTyrThrLeuArgValArgAlaGlnAspGlyGly	480
Db	1381	CCCTTGACTATGAGACGACCAAGAGTACACCTTACGGGTGGACACAGATGGTGGC	1440
QY	481	ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp	500
Db	1441	CGTCCCCCACTCTCTATGCTCTGGCTTGGTGAAGTACAGGTCCTGGATATCAACGAC	1500
QY	501	AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu	520
Db	1501	AATGCCCATCTTCCCTGACAGACCCCTTCCAGGCTACTGTCTCTGAGAGAGCCCTTA	1560
QY	521	GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	1561	GGCTACCTGGTCTCCATGTCAGGCTATCGACGCTGATGCTGGTGAACAATGCCCTG	1620
QY	541	GlyTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560
Db	1621	GAATACCGCTTGTCTGGGGTGGGACATGACTTCCCTTCAACATCAATGACACAGGC	1680
QY	561	TrpIleSerValAlaAlaGlnLeuAspArgGlyGlnValAspPheTyrSerPheGlyVal	580
Db	1681	TGGATCTCTGGCTGCTGTAACCTGACCGGGAGAGATTGATTCTTACAGCTTTGGGGTA	1740
QY	581	GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
Db	1741	GAAGCTCGAGACCATGGCACTCCAGCACTCACTGCTCGGCGAGTCCAGCTGATGTC	1800
QY	601	LeuAspValAlaAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsn	620
Db	1801	CTGGATGTCAAGACACACATCCACTTATCCCAACAGAGTACACATGGCGCTCAT	1860
QY	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis	640
Db	1861	GAGGAGTGCAGCTGTGGGACACAGCGTGTGACCGTGTCACTGTGAGCCGTATGCTCAT	1920
QY	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer	660
Db	1921	AGTGTATACACTTACCAATCACAGTGGCACTGCAAAACCGCTTCTCCATCACACAC	1980
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln	680
Db	1981	CAAGTGGTGGTGGCTGGTATCCCTTGGCTTGGCTGCACTGCACTTGAAGCGGCGAC	2040
QY	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700
Db	2041	TATGTGTTGGCTGTATCCGCTCCGATGACACTCGGACAGACACGACATGTTGGTGTG	2100
QY	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720
Db	2101	AATGTACCGAGCGCAACACCACTGCTCTGTTTCAAGCTCCCACTATACAGGAAT	2160
QY	721	ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740
Db	2161	GTTTAATGAGGACCGGCGGCGGACACACGTTGGTCTGATCAAGCCCAACGATAGAGAC	2220
QY	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2221	ACAGGTGAGATGCGCGCATCACTTCACTATGAGAGACAGCATCCCAAGTTCGCGATC	2280
QY	761	AspAlaAspThrGlyValAlaValThrThrGlnAlaGlnLeuAspTyrGluAspGlnValSer	780
Db	2281	GATGCAAGACCGGGGCTGTCAACACCGCTGACCTGATGATCAAGAAACCAAGTGT	2340
QY	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr	800
Db	2341	TTCACCTCGGCATTACTGCTGGGACCAATGGCATTTCCCAAGATGCCACACCACTTAC	2400
QY	801	LeuGlnIleLeuValAlaAspValAlaAspAsnAlaProGlnPheLeuArgAspSerTyr	820
Db	2401	CTGAGATCTCTGTGAACGACGTGAATGACATGCCCTCAGTCTCTGGAGACCTCTTAC	2460
QY	821	GlnGlySerValTyrGlnAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	2461	CAGGCAAGTGTATGAGAGATGTGCACCTTCACTACAGTCTTCAAGTCTCAACGACT	2520
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly	860
Db	2521	GATCTGATTCCTGACCTTAATGACAGGCTCTTCTACCTTCCAGAGGACCAATGGA	2580
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880
Db	2581	GACGGTGACTTATATTGTTAGTCCACGTACAGGCATGTGCGAACCTTACGAGAGCTGAT	2640
QY	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro	900
Db	2641	CGAGGAACGTGGCCCAATATGTCTTGGCGCATATGACGTGACAAAGGAGATGCCCA	2700
QY	901	AlaArgThrProMetGlyValThrValThrValLeuAspValAlaAspAsnProProVal	920
Db	2701	GGCCGCAACCTATGGAAGTGAAGTCACTGTGTTGATGTGAATGACATCCCTGTTC	2760
QY	921	PheGlnAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAlaVal	940
Db	2761	TTTGAGCAGATAGATTGATGTTGTGGAAAGAACAGGCCATTATGGGCTTACCGTG	2820
QY	941	AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGlnIle	960
Db	2821	GCCCGGTACAGCCACCTGACCCCGATGAAAGCACAAATGCCATATATGACCAAGTT	2880
QY	961	ValGlnGlyAsnIleProGlyValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla	980
Db	2881	GTCGAGGCAACATCCCTGAGGCTTCCAGCTGACACATCTTCCGGGGAAGTGAACACC	2940
QY	981	LeuValAspLeuAspTyrGluAspArgProGlyTyrValLeuValIleGlnAlaThrSer	1000
Db	2941	CTGGTAGACTTATGACTACAGAGACCGGCTGATGATCGTGTATCAGGCGCACGTCA	3000
QY	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAsnPro	1020
Db	3001	GCTCTCTGTGAGCGGGGCTACAGTCCAGTCCGCTCTTGAACCGCAATGACAAACCA	3060
QY	1021	ProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040
Db	3061	CCAAGTCTGGGCACTTGAATCTTTCAACAACATATGTCAACAACTGCTCAAGACGC	3120
QY	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	3121	TTCCCTGGGGGTGCCATTGGCCGATGCTGCCCATGACCTGATATCTCAGATGATCTG	3180
QY	1061	ThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValLeuAsnAlaSerThrGly	1080
Db	3181	ACTTACACCTTGAAGCGGGAAATGAATCAAGCTGTGCTCAATAGCTCCACGCGGT	3240
QY	1081	GluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	3241	GAGCTGAAGCTTAAAGCGCGGACCTGACACAAACCGGCTCTGGAGGCCATCATGACGCT	3300
QY	1101	LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Db	3301	CTGGTGTACAGCGGTGTACACAGGTGACCGCCAGGTGCGGCTGTGATCACTATCATC	3360
QY	1121	ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg	1140
Db	3361	ACCATGAGATGCTTACCAACATCAAGCTGTGGCTTGGAGAGACATTCACCGAGGCGC	3420
QY	1141	PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Db	3421	TTCTGTCAACACATGTATGAGCTCTTATATCAAGGGGTGGCGGACAGCTGGCCACGCA	3480
QY	1161	ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle	1180

D	3481	CCGACACACGTCGTCCTTCAACGTACAGCGGAGCACCCGCGCGCGGCGGCATC	3540	D	4561	GGGCCCCCTCTACTAGCGCGGTGCTTCACTGCCCCGAGAGCTTCCAGTCCG	4620
Q	1181	LeuAnValSerLeuSerValGlyInProProGlyProGlyGlyProProPheLeu	1200	Q	1541	GlnPheValGlyCysMetArgAnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
D	3541	CTCAACGTAGACCTGTGTGGTGGCCAGCGCCGACAGGCGCGGGGGCGGCGCTTCTTG	3600	D	4621	CAGTTGTGGCTGTGACAGCGAACTCTGACGTGGACAGCGGCACTTAACATGGCTTAC	4680
Q	1201	ProSerGlnAspLeuGlnGlnArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer	1220	Q	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIleValysAsnValCysAspSer	1580
D	3601	CCCTCTAGAGACCTGACGAGAGCGCTATACCTCAACCGGAGCGCTGACGCGCATCTCG	3660	D	4681	TTCAATTGCCAACATGGACCGGTGCTGCTGCCCTGCCAAGAAAGACGTGTGACAGC	4740
Q	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGlyProCysGlnAsnTyr	1240	Q	1581	AsnThrCysHisAsnGlyGlyThrCysValAsnGlnITPAspAlaPheSerCysGlyCys	1600
D	3661	GCAAGAGCGGTGCTGCTTCCAGACAACTCTGCGCGGAGCGCTGCGAGAACCTAC	3720	D	4741	AACACTTGCACAAATGGGGGCACTTGTGAAACAGGTGGAGCGCTTACGCTGCGAGTGC	4800
Q	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260	Q	1601	ProLeuGlyPheGlyValysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu	1620
D	3721	ATGCGCTGCGGTGTGTGTGCTGCTGACTCTCTCCGCGCTTCAATGCGCTCTCTCC	3780	D	4801	CCCTTGCGCTTGGGGGCAAGAGCTGCGCCAGAAATGGCCATTCACAGCATTTCTCG	4860
Q	1261	ValLeuPheArgProIleHisProValGlyLeuArgCysArgCysProProGlyPhe	1280	Q	1621	GlySerSerLeuValAlaITPHisGlyLeuSerLeuProIleSerGlnProITPITPLeu	1640
D	3781	GTGCTCTTCCGCGCCATCCACCCGTCGAGAGCGCTGCGCTGCGCGCCGCGCTTC	3840	D	4861	GGCAGAGCTGTGGTCTGGCATGGCTCTGCGTCCCATCTCCAACCTTGTACTTC	4920
Q	1281	ThrGlyAspTyrCysGlnThrGlnValAspLeuCysTyrSerArgProCysGlyProHis	1300	Q	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660
D	3841	ACGGGTACTACTGCGAGACCGAGGTGACCTCTGCTACTCGCGCGCTGTGCGCCAC	3900	D	4921	AGCTCATGTCTCGACCGCGCAGCGCCAGCGGTGTCTGTCTGACGGCCATTCACAGGGGG	4980
Q	1301	GlyArgCysArgSerArgGlnGlyTyrTyrCysLeuCysArgAspGlyTyrThrGly	1320	Q	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGlyThr	1680
D	3901	GGGCGCTGCGCGACCGCGAGGCGGCTTACCTGCTGTGTGTATGGCTTACACGGGT	3960	D	4981	CGAGACACATACCTTACAGCTTACAGAGGCGCCATGATCTAGACCTTGAAGGACAA	5040
Q	1321	GlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly	1340	Q	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspITPHis	1700
D	3961	GAGCACTGTAGAGTGAAGTCTCGCTCAGCGCGTTGCAACCCGGGTGTCTGAAAGATGGG	4020	D	5041	GGGCTTCAAGGCTCTCTCTCGCTGAGCGACGCGCGGCGCAATGACGTGACTGGCAC	5100
Q	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheGlyCysAspCysProSerGlyAspPhe	1360	Q	1701	HisAlaGlnLeuAlaLeuGlyValaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp	1720
D	4021	GGCACCTGTGTCAACCTGTGTGTGGGTTCATGAGCAATGCCATCTGAGAGACTTC	4080	D	5101	CATGACACTGTGCATGTGAAGCAGCGGGGGGCGCTGGCCAGTTCATGTCTGTCAAT	5160
Q	1361	GlnHisProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380	Q	1721	TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisGlySer	1740
D	4081	GAGAGACCTTACTCCAGGTACCAACGCGAGCTTCCCGCCCATCTTCAATCACTTTT	4140	D	5161	TATGGCACAAGAGACAGAGGCAACCTGGGCGCGGCTGCATGTCTGACCTGAGAC	5220
Q	1381	ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrGlyGlnArg	1400	Q	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly	1760
D	4141	CGCGGCTGCGCGCACGCTTCCACTTCACTGCGCCCTCTGTGTCACAAAGAGACGC	4200	D	5221	AACATAACAGTGGCGGAAATACCTGGCCAGCGCGGTGTGCGCTTTCGGGGC	5280
Q	1401	AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlnIleHisAspPheValAlaLeuGln	1420	Q	1761	CysLeuGlnGlyValaArgValSerAspThrProGlnGlyValaAsnSerLeuAspProSer	1780
D	4201	GACGGGTGTGCTGTGTCAATGGGCGTTTCATATGAGAAAGCATGTGTGGCCCTTCAG	4260	D	5281	TGTTTCAGGGGTGTGGGTGAGCCATACGCAAGAGGGGTATTAACGCTGATCCACG	5340
Q	1421	ValIleGlnGlnGlnValaGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer	1440	Q	1781	HisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
D	4261	GTAATCCAGAGAGAGCTCAGCTCACTTCTCTCAGAGGAGTAAACACACGCGTCTCC	4320	D	5341	CATGGGAGAGAGATCAACGTGGAGCAAGGCTGTAGCTTGCAGCTTGTGACTCAAAAC	5400
Q	1441	ProPheValProGlyGlyValaSerAspGlyGlnITPHisITPThrValaGlnLeuTyrTyr	1460	Q	1801	ProCysProAlaAsnSerTyrCysSerAsnAspITPAspSerTyrSerCysSerCysAsp	1820
D	4321	CCATTCGTCGCGGAGAGAGTCAATGATGCGCATAGCGGATAGCGGTGCAATACTAC	4380	D	5401	CCGTGTCTGTAAACGATTAATGCAAGCAAGCTGGGACAGCTATTCCTGACGCTGTGAT	5460
Q	1461	AsnHisProLeuLeuGlnGlnITPThrGlyLeuProGlnGlyProSerGlnGlnValAla	1480	Q	1821	ProGlyTyrThrGlyValAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisGln	1840
D	4381	AATATAGCACTGTGGGTGACAGAGGCTCCACAGGCGCCATACAGAGCAGAAAGGTGGCT	4440	D	5461	CAGGTACTATAGGTATCAACTGTACTAATGTGTGAACCTGAACCGGTGTAGACACAG	5520
Q	1481	ValValThrValaAspGlyCysAspThrGlyValaAlaLeuArgPheGlySerValLeuGly	1500	Q	1841	SerValCysThrArgIleAspSerAlaProHisGlyTyrThrCysGlnCysProProAsn	1860
D	4441	GTGGTGAACCTGTGATGTGTGACACAGAGTGGCTTGGCGCTTGGATCTGTCTGGGC	4500	D	5521	TCTGTGTATACCGGCAAGCCAGTGCCTCCCATAGGTATATACGTGCAAGTGTCCCAAAAT	5580
Q	1501	AsnTyrSerCysAlaIleGlnGlyThrGlnGlyGlySerIleValysSerLeuAspLeuThr	1520	Q	1861	TyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysProArgGlyITPITPITPITP	1880
D	4501	AACACTCTCTGTGTGCTGCGCAAGGACCCAGGGTGGACAAAGAGTCTGTGATCTTACG	4560	D	5581	TACCTTGGGCAATACGTGTGAGACACAGATTGACAGACCTTGTCCCGGTGGCTGGGGA	5640
Q	1521	GlyProLeuLeuGlnGlyValaProAspLeuProGlnSerPheProValArgMetArg	1540	Q	1881	HisProThrCysGlyProCysAsnCysAspValaSerIleGlyPheAspProAspCysAsn	1900
D				D	5641	CATCCACATGTGGCCCATGCACTGTATGTACGAAAGGCTTTGACCAACAGCTCAAC	5700

QY 1901 LyeThSerGlyGluCysHisCysGlyGluAsnHisThrArgProProGlySerProThr 1920  
DB 5701 AAGACAAAGGCGGCGAGGCGCACTGCAAGAGAACACACCGGCCCCAGGACACCCACCC 5760  
QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940  
DB 5761 TGGCTCTGTGTGTACCTGCTACCCCAAGGCTCTTGTCCAGAGTCTGTGACCTTGAGAT 5820  
QY 1941 GlyGlnCysProCysArgProGlyValIleGlyArgGlnCysAspArgCysAspAspPro 1960  
DB 5821 GGGCACTGTCTCAAGCAAGCGGTGTCATCGGGCTGAGTGACCGCTGTGCAACCTT 5880  
QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProAlaIle 1980  
DB 5881 TTGTGAGGTGTCACCAATGAGCTGTGAAGTAATTAAGACGTCGCCACAGGGAAT 5940  
QY 1981 GluAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
DB 5941 GAGGCTGGGATCTGGTGGCCCTTACCCGCTTGGGCTGCTGTGCTGCTCTGCTGCT 6000  
QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGluHisArgGlyTyrLeuProPro 2020  
DB 6001 AAGGCTCTTTTGGACTGCTGTGGCCACTGTGATGAGCAAGGGGGTGGCTCCCCCA 6060  
QY 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGlyPheAlaGluLeu 2040  
DB 6061 AACCTTTCACTGACGTCAGTCATCACCCTTCTCAAGACTGAAGGGCTTCCCTGAGCGGCTA 6120  
QY 2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
DB 6121 CAGCGGAAGTGAAGTCAAGGCTGAGCTAGCGGCGCTCCACAGAGTACCTGCTGCTGCGC 6180  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIleArgGlnLeu 2080  
DB 6181 AAGCGCACCGACGACACACTGGCTACTTCGGGACGAGGAGTCAAGTGGCTTACCAAGCTG 6240  
QY 2081 AlaThrArgLeuLeuAlaHisGlnSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
DB 6241 GCCACCGGCTGTGGCCCAAGAGACACCGAGGGGCTTGTGGCTGTGCTGCAACAG 6300  
QY 2101 AspValHisPheThrGlnLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
DB 6301 GAGGTGCACTTCACTGAGAAATCTGCTGGGGTGGGAGCGCTCTTGACACAGCAAC 6360  
QY 2121 LysArgHisThrGlyLeuLeuIleGlnGlnThrGlyGlyThrAlaThrLeuLeuGlnHis 2140  
DB 6361 AAGCGGCACTGGGAGCTGATCCAGACAGAGGGTGGCAAGCGCTGGCTCCAGCAC 6420  
QY 2141 TyrGluAlaThrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160  
DB 6421 TATGAGGCTTACGCGCAAGTCCCTGGCCAGAACATGGGCAACCTTAAAGCCCCCTTC 6480  
QY 2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspLysGlyAsnPheAla 2180  
DB 6481 ACCATGTACCGCCCAACATTGTCACTCCGTAAGTCCGCTTGACAAAGGAACTTTGCT 6540  
QY 2181 GlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlyGlnProProAspLeuGlnThr 2200  
DB 6541 GGGGCGCAAGCTGCCGCTACGAGGCGCTGGGAGGAGCAAGCCCCGGAACCTTGAGACA 6600  
QY 2201 ThrValIleLeuProGlnSerValPheArgGluThrProProValIleArgProAlaGly 2220  
DB 6601 AAGATCATTTCTGCTGAGTCTGTCTTCAAGAGAGCCCCCGGTGTACGGCCCGAGGCG 6660  
QY 2221 ProGlyGluAlaGlnGluProGlyGlnLeuAlaArgArgGlnArgArgHisPProGlyLeu 2240  
DB 6661 CCGGAGAGGCGCCAGGAGCGAGGAGCTGGACCGGCGACAGGAGCGGACCCGAGGAGCTG 6720  
QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuPro 2260  
DB 6721 ACCCAGGGTGAAGCTGTGGCGAGCGTCATCATTCAGCACCTGTGGCCGGGCTACTGCT 6780

QY 2261 HisAsnTyrAspProAspArgLysSerLeuArgValProLysArgProIleIleAsnThr 2280  
DB 6781 CATTAATCTAGACCTTGACAAAGGCGAGCTTGAGAGTCCCAAGCGCCGATCATTAACACA 6840  
QY 2281 ProValIleSerIleSerValHisAspAspGlnGluLeuLeuProArgAlaLeuAspLys 2300  
DB 6841 CCGGTGTGAGATCAAGGTCTCAT 6900  
QY 2301 ProValIleGlnPheArgLeuLeuGlnThrGlnGlyArgThrLysAspProIleCysVal 2320  
DB 6901 CCGTCAAGGTCAAGTTCGCTGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6960  
QY 2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgLysGlu 2340  
DB 6961 TTCTGAACCATTCATCTGCTGTGTCAGTGGCAAGGTGTGCTGTGGCCAGAGGCTGTGAA 7020  
QY 2341 ValValPheArgAsnGlnSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
DB 7021 GTGCTTCCGCAATGAGAGGCACTGACGTGCCAGTGCACACATAGAGACTTCGCT 7080  
QY 2361 ValLeuMetAspValSerArgArgGluAsnGlyGlnIleLeuProLysThrLeuThr 2380  
DB 7081 GTGCTCATGAGAGTTCCTCGGCGGAGGAATGGGAGATCTTGCCACTGAAGCACTGACA 7140  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400  
DB 7141 TACGTGGCTTGAAGGTCACTGTGCTGCCCTTCTGCTCACTTCTTCTTCTTCTTCTTCTC 7200  
QY 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgGluAsnLeuThrAlaAlaLeuGly 2420  
DB 7201 TTGCTTATCTGCGCTTCCAAACCAACCGGATCCAGCTTACCTTACAGCTGCTGCGGC 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyTyrIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
DB 7261 CTGGCTCAGCTGAGTTCCTCTCGGGATTCACAGGCTGACCTGCTTTGGCTGACACA 7320  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuLeuGlu 2460  
DB 7321 GTCATTCGATCTGTGCTGACCTTCTGTACCTGTGACCTTCTTCTGAGCTGTGTGAG 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480  
DB 7381 GCTTGGACCTGTACCGGGCACTCAAGGTGCCGATGTCCACACCGGCCCCATGTGCGC 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
DB 7441 TTCTACTACATGCTGGGCTGGGCGCTGCTGCTTCAATCAAGGCTAGCGTGGGCTG 7500  
QY 2501 AspProGlnGlyTyrGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuIle 2520  
DB 7501 GACCCCGAGGCTACGGGAGACCTTGACTTGTGCTGCTTCATCTATGACACGCTCATC 7560  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
DB 7561 TGGAGTTTGTGTGGCCGAGTGGCTTTCGCTGAGTATGTCTTCTGTACATCTG 7620  
QY 2541 AlaAlaArgAlaSerCysAlaIleGlnArgGlnGlyPheGlnLysGlyProValSer 2560  
DB 7621 GCGGCGCGGCGCTGCTGCTGCTGCCAGGCGGAGGCTTGAGAAAGGCTGTCTG 7680  
QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
DB 7681 GAGCTGACGCTCTTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7740  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisThrLeuPheAlaThrCysAsnCysIle 2600  
DB 7741 CTCTCTGTCAACAGACACACCTCTCTTCACTACCTTGTGCTACCTGCAATTTGATC 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerLysGluValArgLysAlaLeu 2620  
DB 7801 CAGGCGCCCTTCACTTCTCTCTATGTGTGTGTGCAAGGAGGCTCCGCGGAAAGCACTC 7860  
QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640



```

Db      ||| 7861 AAGCTTGCTGCGAGCGCAGACCCCTGACCTGCTGACCAAGTCCACCTCG 7920
Qy      ||| 2641 ThrSerSerTyranCysProSerProTyraAlaAspGlyArgLeuTyrglnProTyrgly 2660
Db      ||| 7921 ACCTGCTCTCACTACCTGCCCCACCTTACGCAATGGGGGCTGTACCAAGCCCTTACGA 7980
Qy      ||| 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyValSerSerGlnProSerTyrgly 2680
Db      ||| 7981 GACTCGCGCGCTCTCTGCAACAGCTCGCTCGGCAAGAGTACGACCCAGCTACATC 8040
Qy      ||| 2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyValGlnGlyProProGlyLeuGly 2700
Db      ||| 8041 CCTCTTGTGTGAGGAGGAGTCCGCACTGAACCTTGCGCAAGGCGCCCTGCTCGGCG 8100
Qy      ||| 2701 AspProGlySerLeuPheLeuGlnGlnAspGlnGlnHisAspProAspThrAspSer 2720
Db      ||| 8101 GATCCAGGAGGAGCTTCTTCTGAGAGGTCAAGACCAAGACGATATCTTACACGAGCTCC 8160
Qy      ||| 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyraAlaSerThrHisSerSer 2740
Db      ||| 8161 GACAGTGAAGCTGCTTGTAGAGAGCAGACAGAGTGGCTCTTATGCTTACCCACTCATCA 8220
Qy      ||| 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGly 2760
Db      ||| 8221 GACAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280
Qy      ||| 2761 TrpAspSerLeuLeuGlnProGlyAlaGlnArgLeuProLeuHisSerThrProTyraAsp 2780
Db      ||| 8281 TGGGATAGCTGCTGTGGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8340
Qy      ||| 2781 GlyGlyProGlyProGlyValAlaProTyraProGlyValAspPheGlyThrThrAlaValGln 2800
Db      ||| 8341 GGGGGCCCAAGGGGCTGCAAGGCGCCCTGCGCAGAGACTTTGAGACCAAGCAAGAGAG 8400
Qy      ||| 2801 SerSerGlyValAsnGlyAlaProGlnGlnArgLeuArgGlnAsnGlyValAspAlaLeuSerArg 2820
Db      ||| 8401 AGTAGTGGCAACGGGGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8460
Qy      ||| 2821 GlnGlySerLeuGlnProLeuProGlySerSerAlaGlnProHisGlyIleLeuGly 2840
Db      ||| 8461 GAGGGGCTCCCTAAGGCCCTTCCAGAGCTCTTCCAGAGCTCAAAAGGAGCACTTAAAG 8520
Qy      ||| 2841 LysGlyCysLeuProThrIleSerGlyValSerSerLeuLeuArgLeuProLeuGlnGln 2860
Db      ||| 8521 AAGAAATGTCTGCCCAACATCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8580
Qy      ||| 2861 CysThrGlySerSerArgGlySerSerAlaSerGlnGlySerArgGlyValGlyProProPro 2880
Db      ||| 8581 TGCACAGGGCTTCTCCGGGGCTCTCCGCTGATGAGGGAGAGAGAGAGAGAGAGAGAGAG 8640
Qy      ||| 2881 ArgProProProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2900
Db      ||| 8641 CGCCCAACCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8700
Qy      ||| 2901 SerIleValAlaGlyThrValAlaGlnAspSerSerGlySerGlnPheLeuPheLeu 2920
Db      ||| 8701 AGCATCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8760
Qy      ||| 2921 PheLeuHis 2923
Db      ||| 8761 TTCCTGAT 8769

```

RESULT 2  
AB242868  
ID AB242868 standard; DNA; 10531 BP.  
XX

AC AB242868;  
XX

DT 04-MAR-2003 (first entry)  
XX

DE Human GPCR CELSR2 nucleotide SEQ ID NO:523.  
XX

KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
KW G protein-coupled receptor modulator; antibody; immune-related disease;  
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
KW immunological-related cell proliferative disease; autoimmune disease;  
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
KW gastritis; anxiety; depression; Parkinson's disease; multiple sclerosis; pain;  
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;  
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
KW ulcer; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200261087-A2.  
XX  
PD 08-AUG-2002.  
XX  
PF 19-DEC-2001; 2001WO-US50107.  
XX  
PR 19-DEC-2000; 2000US-257144P.  
XX  
PA (LIFE-) LIFESPAN BIOSCIENCES INC.  
XX  
PI Burner GC, Roush CL, Brown JP;  
XX  
XX WPI; 2003-046718/04.  
XX  
XX P-PSDB; ABP82018.  
XX  
PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
PT (GPCR), useful for diagnosing and designing drugs for treating  
PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
PT cancer or autoimmune diseases -  
XX  
XX  
PS Disclosure; Fig 1; 523pp; English.  
XX  
CC The present invention describes antigenic peptides (I) comprising:  
CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
CC acids. Also described: (1) an assay for the detection of a particular  
CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity  
CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
CC an antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related disease, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention.  
XX  
XX Sequence 10531 BP; 2038 A; 3353 C; 2998 G; 2142 T; 0 other;  
SQ

#### Alignment Scores:

Pred. No.:	0	Length:	10531
Score:	2923.00	Matches:	2923
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	25	Gaps:	0

US-09-916-849A-3 (1-2923) X AB242868 (1-10531)

```
QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20
DB 63 ATGCGAAGCCGGCCACCGGCGTCCCTCCCAAGCCGCGCCCTGCTGCTG 122
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAAspGlnValGlyProCysArgSerLeu 40
DB 123 TTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182
QY 41 GlySerArgGlyValArgGlySerSerGlyAcyAalAProMetGlyTyrLeuCyAspSer 60
DB 183 GGGTTCAGGGGAGCGAGGCTCTTGGGGGCTGCGCCCAATGGCTGGCTGCTCATCC 242
QY 61 SerAlaSerLeuLeuTyrThrLeuTyrThrSerArgCysArgAspAlaGlyTyrThrLeu 80
DB 243 TCAAGGTGAACCTGTGCTTACACAGCGCTCAGAGGAGCGAGGAGCTGAGCTGACT 302
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlySerGlyValHis 100
DB 303 GGCCACCTGGTACCCCAACGATGCGCTGAGGGTTGGTGTCCGAATCCGAGGCCAT 362
QY 101 IleProLeuProProAlaProGlyLeuGlyCysProTyrSerCysArgLeuLeuGlyTyrLeu 120
DB 363 ATTCCCTACCAACAGCTCTGTAAGGCTGCGCTGAGCTGTGCTCTGGGCAATTGGA 422
QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlyLeuHisProCysLeuValA 140
DB 423 GGGCACTTTCCCAACAGGAGCTCACCTGCGCCGAGAGCACCCGTGCTTAAAGCT 482
QY 141 ProArgLeuArgCysGlnSerCysLeuLeuAlaGlnAlaProGlyLeuArgAlaGlyVal 160
DB 483 CCAAGGCTCAATGCGACTCTGCAAGCTGGGCAAGGCCCGCGGCTCAGGCGAGGGA 542
QY 161 ArgSerProGlnGlySerLeuGlyValArgArgValAsnThrAlaProGln 180
DB 543 AGGTCAACCAAGAGTCCCTGGGTGGCGTGGAAAGAAATGTAAATACAGCCCCCA 602
QY 181 PheGlnProProSerTyrGlnAlaThrValProGlyAsnGlnProAlaGlyTyrThrProVal 200
DB 603 TTCAGGCCCCCAGCTACAGGAGCAAGTGCAGGAAACAGCCAGGAGGACCCCTGT 662
QY 201 AlaSerLeuArgAlaIleAspProAspGlnGlyValArgLeuGlyTyrThrMet 220
DB 663 GCATTCCTGAGAGGCGATGACCCCGGACGAGGCTGAGGAGCTGACTGAGTACCAATG 722
QY 221 AspAlaIleuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
DB 723 GATGCCCTCTTGATAGCGCGCTCCAAACAGTTCTTCTCCCTGGAACCGACTGCTGCA 782
QY 241 ValThrThrAlaGlyLeuLeuAspArgGlyThrTyrSerThrHisValPheArgValThr 260
DB 783 GTAAACCAAGCCGAGAGAGCTGATCGTGAACCAAGACCAACCGTCTTCAAGGATCAG 842
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280
DB 843 GCGAGAGACCAACGGCATGCCCCGACGAAAGTGCCTGCTTACATCCATTTGGTAACT 902
QY 281 AspThrAsnAspHisAspProValPheGlyLeuGlnGlyTyrGlySerLeuArgGln 300
DB 903 GACACCAATGACCATGACCTGTGTGAGCAGCAGGAGTCAAGAGAGGCTCAGGGAG 962
QY 301 AsnLeuGlyValGlyTyrGlyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320
DB 963 AACCTGGAGGTGCTATGAGGTCTCACTGTCAGGGCCAGGATGTGAGTCCCTCC 1022
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlyValSerGlyValSerProSerGlyValPhe 340
DB 1023 AATTCGCAATATTTCTGTAACGCTGCTGAGAGGGCTCTGGGGGCAAGCCCTCTGAAGTCTTT 1082
QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlyVal 360
DB 1083 GAGATGACCTCTGCTGAGGTATCCGAACCGTGGCCCTGTGATCGGGAAGAGGTG 1142
QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380
DB 1143 GAATCTTACACAGCTGACGATGAGGAGCAAGTGAACAGGGTCCGGGCTCTCGAGAT 1202
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400
DB 1203 ACCACAGCCGCTGTTTCTCTTCTGTGTGAGATGACAAATGAATGCCCCCAAGTTAAT 1262
QY 401 GluIleAspGlyValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
DB 1263 GAGAAAGGCTATGTGTGTCAGGTGAGGAGGATGTGACTCCAGGGGGCCCAAGTCTCCGA 1322
QY 421 ValThrAlaSerAspAspAspGlySerAsnAlaValValHisTyrSerTyrMetSer 440
DB 1323 GTCAAGCTCTGATTCAGACAGAGGAGCAATCCGTGTGTGACTATGACATCATGAT 1382
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaAspValValSer 460
DB 1383 GGCAATGCTGGGAGACGTTTATCTGATGCCCCAGACTGAGCTGTGATGTGTGAGC 1442
QY 461 ProLeuAspTyrGlyLeuThrTyrGlyValTyrThrLeuArgValArgAlaGlnAspGlyVal 480
DB 1443 CCTCTTGAATATGAGACAGACCAAGAGATACACCTTACGGGTGCGAGCAGAGATGTGGC 1502
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
DB 1503 CGTCCCCACTCTCTAATGTCTGTGCTGTGTGACATACAGGTCTGTGATCAACGAC 1562
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlySerValProLeu 520
DB 1563 AATGCCCTTCTTGTGACGACCCCTTTCAGGCTACTGTGCGAGAGGTCCCTTA 1622
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
DB 1623 GGCTACTGTGTTCCATGATCCAGGCTATGAGCTGATCTGTGTGACATGCCGCTG 1682
QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly 560
DB 1683 GAATACCGCTTCTGTGGGGGAGCAGATCTTCCCTTCAACATCAATGAGCAGAGC 1742
QY 561 TrpIleSerValAlaAlaGlyLeuLeuAspArgGlyValAspPheTyrSerPheGlyVal 580
DB 1743 TGGATCTCTGTGCTGCTGAACCTGAGCCGAGAGAGATTAATTTCTACAGCTTTGGGATA 1802
QY 581 GluAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValThrVal 600
DB 1803 GAAGCTGAGACCATGCACTCCAGGACTCATGCGCTCGGCAAGTGTCAAGGTACTGTC 1862
QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsn 620
DB 1863 CTGATGTCAACGACAAACATCCAACTTACCAACCAAGATGACAGTGGCGCTCAAT 1922
QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspAspAlaHis 640
DB 1923 GAGGATACACTGTGGGACCAACGCTGTGACCGTGTCACTGTGACCGTATGCTCAT 1982
QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
DB 1983 AGTGTATCACTTACCAAGATCAACAGTGGCAATCTGAAACCGCTTCTCATCACAGC 2042
QY 661 GlnSerGlyGlyValLeuValSerLeuAlaLeuProLeuAspTyrIleValGlnValGln 680
DB 2043 CAAGTGTGTGTGGGTGTGATCCCTTGCCTTCCACTGACATCAAACTTGGAGCGGAG 2102
QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
DB 2103 TATGTGTGCTGTATACGCTCCGATGCACTCGGAGACACGGACAGATTTGTGTG 2162
QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
DB 2163 AATGTCAACGAGCCAAACCAACCATGCTGTCTTCAAGGCTCCCACTATACAGTGAAT 2222
QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
```

Db	2223	GTATATAGAGACCGGGCCGGCAGGACCAACGGTGGTGTGATCAAGGCCACCGATGAGAAC	2282
Qy	741	ThrgIygiuamnlaaargilleThrTyRphenetgiuaspserilleProginPheargIle	760
Db	2283	ACAGGTGGAATGCCCGCATCACTTCACTTGAAGGACACACATCCCCCACTCCGCAATC	2342
Qy	761	AspAlaaspThrgIyAlaValThrThrcGlnIaagiueuaspTyrgIuaspGlnValaser	780
Db	2343	GATCAGACACGGGGGGCTGTCAACCCAGGCTAGCTGACTACAGAGACCAAGTGTCT	2402
Qy	781	TyThrIleuAlaIleThrAlaargAspansgIyIleProginIySeraspThrThrTyR	800
Db	2403	TACACCCCTGGCCATTACTGTCTCGGGACAATGGCATTTCCCAAGAGTCCGACCACTTAC	2462
Qy	801	LeuGlnIleuValAspAspValAspAspAspAlaProGlnPheLeuArgAspSerTyR	820
Db	2463	CTGGAGATCTCGGTGAGACGAGTGATACAAATGCCCTCAAGTTCCTCGAGACCTCTTAC	2522
Qy	821	GlnIySerasValTyrgIuaspValProProPheThrSerasValLeuGlnIleSeralAthr	840
Db	2523	CAGGGCAGTGTCTAATGAGATGTGCACCTTCACTACGCTCTGCAATCTCAGACCACT	2582
Qy	841	AspArgaspSergIyLeuAspansgIyAryValPheTyThrPheGlnIyGlyAspaspGly	860
Db	2583	GATCGTGAATCTCGACCTTAATGAGCGGCTCTTCAACCTTCAAGGAGCGACGATGGA	2642
Qy	861	AspGlyAspPheIleValIySerasThrsersgIyIleValArgThrLeuArgArgLeuasp	880
Db	2643	GACGTGACTTATATGTGTGAGTCCACGTCAGGCATCTGTCGACAGCTACGAGCTGGAT	2702
Qy	881	ArgGluAspValAlaGlnTyValIleuArgAlaTyAlaValAspLygIymeProPro	900
Db	2703	CGAGAGAACGTGGCCCACTATGTCTTGGCGGCATATGACGTGACGACAGGGATGCCCA	2762
Qy	901	AlaArgThrProMetGluValThrValThrValIleuAspValAspAspProProAla	920
Db	2763	GCCCGCACCTTAATGAGAGTGAACAATCACTGTGTGATGATGAAATGACAAATCCCCGTCTC	2822
Qy	921	PheGlnIuAspGluPheaspValPheValGlnIuAspSerProIleGlyLeuAlaVal	940
Db	2823	TTTGAAGCAGATGAGTTGATGTGTGTTGTGGAAGGACAGGCCCATTTGGGCTAGCCGTG	2882
Qy	941	AlaArgValIThrAlaThrAspProAspGlnGlyIThrAsnAlaGlnIleMetTyGlnIle	960
Db	2883	GCCCGGATCAACGACCTGACCCCGATAAAGGCACCAATGCCCAATTAATATGACAGATT	2942
Qy	961	ValGlnGlyAsnIleProGlnValPheGlnIleuAspIlePheSergIyGluThrAla	980
Db	2943	GTGAGGGGCACATCTCCAGAGTCTTCAAGCTGACACCTTCTCCGGGGAAGCTACAGCC	3002
Qy	981	LeuValAspLeuaspTyrgIuaspArgProGlnTyValIleuValIleGlnAlaThrser	1000
Db	3003	CTGTGAGACTTAAGCTACAGAGACCGGGCTGATCGTCTGTGTATCAAGGCCACGTCA	3062
Qy	1001	AlaProIleuValIleSerasAlaThrValIAsnIleValArgLeuAspAspArgAspAspPro	1020
Db	3063	GCTCTCTGTGTGAGCCGGGCTACATCACTCCGCTCTCTTGACCGCAATGACACCCA	3122
Qy	1021	ProValIleuGlyAsnPheGlnIleLeuPheAsnAntyValIThrAsnArgSerasSer	1040
Db	3123	CCAGGTGTGGGACCTTGAAGATCTTTTCAACACATATGTCAACCAATCGCTCAAGCAGC	3182
Qy	1041	PheProGlyGlyValIleGlyAryValProIAsnIleAspProAspIleSerasPserIleu	1060
Db	3183	TTCCCTGTGGGGGCGCAATGTGGCGAATCACTGCCCACATGACCCGATATCTCAATATGTGTG	3242
Qy	1061	ThrTyRserPheGlnuArgGlyAsnGlyIleuSerasIleuValIleuLeuAsnAlaserThrgIy	1080
Db	3243	ACTTACAGCTTTGAGCGGGGAAATGAACCTCAGCTGTCTGTCTCAATGGCTCCAGGGGT	3302
Qy	1081	GluLeuLyIleuSerasArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerasVal	11000

Dd	3303	GAGCTGAAGGCTAAGCGCGCACTGGACAACAACCGCGCTCTGGAGGCGCATATGAGGCTG	3362
Qy	1101	LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Dd	3363	CTGGATGTCAGACGGCGCTGACACAGAGGTACACGGCCACAGCGCGCGCTGTGGACATCATC	3422
Qy	1121	ThrAspGluMetLeuThrHisSerIleGThrLeuArgGluIuAspMetSerProGluArg	1140
Dd	3423	ACCGATAGATGCTCACCCACAGCATACCGCTGGCGCTGGAGAGCATGTACACCCGAGCGC	3482
Qy	1141	PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Dd	3483	TTCTCTGCACCACTGCTAGGCGCTTTCATCCAGCGCGGTGGCCCGGCACGCTGGCCACCCCA	3542
Qy	1161	ProAspHisValValAlaPheAsnValGlnArgAspThrAspHisIleProGlyValHisIle	1180
Dd	3543	CCGAGACCACGTGGTGGCTTTCACAGTACAGGGGAGACCCAGCCGCCCGGGGGGACACATC	3602
Qy	1181	LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu	1200
Dd	3603	CTCAACGTGAGCGTGTGGTGGGCAACCGGACAGGGCCGGGGGGCGGGCCCTTCTCTG	3662
Qy	1201	ProSerGluAspLeuGlnGluArgLeuTyrlLeuAsnArgSerLeuThrAlaIleSer	1220
Dd	3663	CCCTCTAGGACCTCGACGAGCGGCTATACCTCAACCGAGCCCTGACGCGGCATCTCG	3722
Qy	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgIuProCysGluAsnTyx	1240
Dd	3723	GCAAGGCGCGTGGCGCTTTCAGACGACATCTGCTGGGGAGACCTTCGAGAACTAC	3782
Qy	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260
Dd	3783	ATGGCGTGGCTGTGCGGTGGCTGGCTTGCATCTCCCGCGCCCTTCATCGCTCTCTCC	3842
Qy	1261	ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGlyPhe	1280
Dd	3843	GTGCTCTTCGGGCCCATTCACCCCGTGGAGGGCTGGCGCTGCGCTGCGCCGGCTTC	3902
Qy	1281	ThrGlyAspTyxArgGluThrGluValAspLeuCysTyxSerArgProCysGlyProHis	1300
Dd	3903	ACGGGTGACTACTCGAGACCGAGGTGACCTTGTGTACTCGCGCGCTGGCGCCCCAC	3962
Qy	1301	GlyArgCysArgSerArgGluGlyGlyTyxThrCysLeuCysArgAspGlyTyxThrGly	1320
Dd	3963	GGGGGCTGGCGAGCCCGAGGGGGGTACACCTGCTGTGCTGTAAGGCTTACACGGGT	4022
Qy	1321	GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIysAsnGly	1340
Dd	4023	GAGCACTGTAGGTGAAGTGTGGCTGCTCAAGGCGGTTGACCCCGGGTGTCTGGACAAATGGG	4082
Qy	1341	GlyThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysProSerGlyAspPhe	1360
Dd	4083	GGACCTGTGTGCAACCTGTGGTGGGGGTTTCAAGGTGCATTGCCATCTGGAGACTTC	4142
Qy	1361	GluIysProTyxArgGlnValIleThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380
Dd	4143	GAGAAAGCTTACTGCCAGGTGACCAACCGGACGTTCCCGGCCACTCTTCATACACTTT	4202
Qy	1381	ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIysGluArg	1400
Dd	4203	CGCGGCTGGCGCACGGCTTTCACCTTCAACCTTGGCCCTCTGTGTGGCACAAAGACGCG	4262
Qy	1401	AspGlyLeuLeuLeuTyxArgGlyArgPheAsnGlyIysHisAspPheValAlaLeuGlu	1420
Dd	4263	GACGGTGTGCTGTGTGCAATGGCGGTTTCATATAGAAACATGACTTTGTGGCCCTGGAG	4322
Qy	1421	ValIleGlnGlnValAlaGlnLeuThrPheSerAlaGlyGluSerThrThrValSer	1440
Dd	4323	GTGATTCACAGGACGATTCAGCTCACTTCTCTGACGGGAGGTCAACAACAACGAGTCTCC	4382
Qy	1441	ProPheValProGlyValIleSerAspGlyGlyThrPheThrValGlnLeuValTyxTyx	1460
Dd	4383	CAATTTCGGCCGGAGAGATGATGATGAGCGCAAGTGGCAATACGGTGCAGCTGAATATCTAC	4442

QY 1461 AsnlypProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
Db 4443 AATAAGCCACTGTGGGTCAAGCAGGGCTCCACAGGGCCCATCAGACAGAGGTGGCT 4502  
QY 1481 ValValThrValAsnGlyCyAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4503 GTGGGACCGGTGGAGTGGCTGTGACACAGAGTGGCTTGGCTTGGATCTGTCTGGGC 4562  
QY 1501 AsnTySerCyAlaAlaGlnGlyThrGlnGlyGlySerTyLeuSerLeuAspLeuThr 1520  
Db 4563 AACTACTCTGTGCTGCCACAGGGCACCCAGGGTGGACAGAGATCTCTGATTTGACG 4622  
QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGlnSerPheProValArgMetArg 1540  
Db 4623 GGGCCCCCTGACTAGCGGGGTGGCTGACCTGCCAGAGCTTCCAGTCCGAATGGCG 4682  
QY 1541 GlnPheValGlyCyMetArgAsnLeuGlnValAspSerArgHisAlaAspMetAlaAsp 1560  
Db 4683 CAGTTCGGGGCTGCATGCGGAACTGAGGTGACACCGGCACATAGACATGGCTGAC 4742  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCyProAlaTyLeuAsnValCyAspSer 1580  
Db 4743 TTCAATGCCAACATAGGACCGGTGCTGGCTGCTGCCAGAGAAAGATGTGTGACAGC 4802  
QY 1581 AsnThrCyHisAsnGlyGlyThrCyValAsnGlnTTPAspAlaPheSerCyAspGlyCyS 1600  
Db 4803 AACACTGCCACATAGGGGGACCTTGGCTGAACCACTGGAGAGCGGTTCAGCTGCGAGTGC 4862  
QY 1601 ProLeuGlyPheGlyGlyTyLeuSerCyAlaGlnGlyMetAlaAsnProGlnHisPheLeu 1620  
Db 4863 CCCCCTGGGCTTTGGGGGCAAGAGCTGTCGCCAGAAATGGCCATCCACACACTTCTGTG 4922  
QY 1621 GlySerSerLeuValAlaTTPHisGlyLeuSerLeuProIleSerGlnProTTPTyLeu 1640  
Db 4923 GCGACGACCTCGTGGCTGGCATGGCTTCTGCTGCCATCTCCCAACCTTGGTACTTC 4982  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4983 AGCTCATGTTCCGCACCGCCGCGACGGCTGCTGACAGGCATCACAGGGGG 5042  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlnGlnHisValMetLeuSerValGlnGlyThr 1680  
Db 5043 GCGACGACCATCACCCTACAGCTACAGAGGGCCACGTGATGTGAGGTGAGGGCA 5102  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTTPHis 1700  
Db 5103 GGGCTTCAGGCTCTCTCTCTCGTGTGAGGCCAGGCCCGGCCAATGACGGTGACTGGCAC 5162  
QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
Db 5163 CATGCACAGCTGGCACTGGAGACGAGGGGGGCTTGGCCATGCCATTCTGCTTCGAT 5222  
QY 1721 TyrGlyGlnGlnAlaArgAlaGlnGlyAsnLeuGlyProAlaGlnHisGlyLeuHisLeuSer 1740  
Db 5223 TATGGGCGAGAGAGCGAGAGGGCAACTGGGCCCCCGGCTGCATGTGCTCACCTGAC 5282  
QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAspGlyPheArgGly 1760  
Db 5283 AACATACAGTGGGGGGAATCTTGGCCAGCCGGGTGTGGCCCTTGGGCTTTCGGGGC 5342  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer 1780  
Db 5343 TGTTTGACGGGTGTGCGGGTGAAGCATGCCCAAGGGGGTTAACAGCTGATCCCGAC 5402  
QY 1781 HisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCyAspSerAsn 1800  
Db 5403 CATGGGAGAGCATCAAGGTGAGCAAGGCTGTAGCTGTGACCTTGTGACCTCAAC 5462  
QY 1801 ProCyProAlaAsnSerTyCySerAsnAspTTPAspSerTySerCySerCyAsp 1820  
Db 5463 CCGTGTCTCTGTACAGCTATTGACGACACGACTGGGACACTATCTCTGACGCTGTGAT 5522

QY 1821 ProGlyTyTyGlyAspAsnCyThrAsnValCyAspLeuAsnProCyGlnHisGln 1840  
Db 5523 CCAAGTTACTATGGTGAACAAGTACTAATGTGTGACCTGAACCCGTGTGACACAG 5582  
QY 1841 SerValCyThrArgTyProSerAlaProHisGlyTyThrCyGlnGlyProProAsn 1860  
Db 5583 TCTGTGTACCCCGCAAGCCAGTGGCCCCCATGGCTATGCTGCGATGTGCCCCCAAT 5642  
QY 1861 TyLeuGlyProTyTyGlyGlnThrArgIleAspGlnProCyProArgGlyTTPGly 1880  
Db 5643 TACTTGGGCAATCTGTGAGACCAAGATTGACCAAGCTTGTCCCCCGTGGTGGGGA 5702  
QY 1881 HisProThrCyGlyProCyAsnCyAsnValSerTyGlyPheAspProAspCyAsn 1900  
Db 5703 CATCCCAATGTGGCCATCACTGATGATGTGACAAAGGCTTGTGACCCAGCTGCAC 5762  
QY 1901 TyThrSerGlyGlyCyHisGlyGlyGlnAsnHisGlyArgProProGlyTySerProThr 1920  
Db 5763 AAGCAAGCGCGAGTCCACTGCAAGAGAAACACTACCGGCCCAAGGACGCCAC 5822  
QY 1921 CysLeuLeuCyAspCyTyTyProThrGlySerLeuSerArgValCyAspProGlnAsp 1940  
Db 5823 TGCCTTGTGTGACTGTACCCCAAGGCTCTTGTCCAGAGCTGTGACCTGAGAT 5882  
QY 1941 GlyGlnCyProCyTyProGlyValIleGlyArgGlnCyAspArgCyAspAsnPro 1960  
Db 5883 GGCAGAGTGCATGACAGACAGAGGTGATCAGGGGTGAGTGAACCTGTGACAACTT 5942  
QY 1961 PheAlaGlyValThrThrAsnGlyCyGlnValAsnTyThrAspSerCyProArgAlaIle 1980  
Db 5943 TTGTCTAAGGTGCACACAAATGGCTGTGAAGTGAATATGACAGCTGCCACAGGCATTT 6002  
QY 1981 GlyAlaGlyIleTTPTyProArgThrArgPheGlyLeuProAlaAlaAlaProCyPro 2000  
Db 6003 GAGGCTGGATCTGTGTGGCCCCCGTACCCGCTTGGGCTGCTGTCTCTCTGCTCC 6062  
QY 2001 TyGlySerPheGlyThrAlaValArgHisCyAspGlnHisArgGlyTTPLeuProPro 2020  
Db 6063 AAAGGCTCTTGTGGACTGCTGTGGCCCACTGTGATGACACAGGGGTGGCTCCCCCA 6122  
QY 2021 AsnLeuPheAsnCyThrSerIleThrPheSerGlnLeuTyGlyPheAlaGlnArgLeu 2040  
Db 6123 AACCTTCAACTGCACAGTCACTCACTTCTCAGAACTGAAGGGCTTGGTACGGCTA 6182  
QY 2041 GlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnAlaLeuLeuArg 2060  
Db 6183 CAGCGAAATAGTCAAGGCTAGACTCAGGGCGCTCCAGAGCTAGGCTGCTCTGCGC 6242  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTyThrPheGlySerAspValIleValArgGlnLeu 2080  
Db 6243 AACGCCACGACGACACAGCTGCTACTTCCGACAGCAAGCTCAAGGTGGCTTACAGCTG 6302  
QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6303 GCCACGGGCTGTGGGCCCAAGAGACCCACAGGGGGCTTGGGCTGTGCGCACAG 6362  
QY 2101 AspValHisPheThrGlnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6363 GAGCTGACCTTCACTAGAGATCTGCTCGGGTGGGACGCCCTTCTGACACAGCCAC 6422  
QY 2121 TyArgHisTTPGlyLeuLeuIleGlnGlnThrGlnGlyGlyThrAlaTTPLeuLeuGlnHis 2140  
Db 6423 AAGCGGCACTGGAGGCTGATCCAGACAGAGGTGGACCCCTGCGTGCACAC 6482  
QY 2141 TyrGlnAlaTyThrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyThrLeuSerProPhe 2160  
Db 6483 TATGAGGCTTACGCAAGTGCCTGGCCCAAGACATGGGGCACTTACCTTAAGCCCTTTC 6542  
QY 2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspTyGlyAspPheAla 2180  
Db 6543 ACATGCTGACGGCCAAATGTGATCTCGTGTGGCTTGGACAAAGGAATTTGCT 6602  
QY 2181 GlyAlaTyLeuProArgTyGlyAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200

6603 GGGGCGAAGCTGGCCGCTACAGAGCCCTGGCGGAGAGAGCCCGGAGCCTTGAAGCA 6662  
2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProIleagly 2220  
6663 ACGATATCTGCTGATGCTGTCTTCAAGAGAGCCCGGCTGGTCAAGGCCCGGAGGCG 6722  
2221 ProGlyValIleaglnGluProGluGluLeuValArgArgGlnArgArgIleProGluLeu 2240  
6723 CCGGAGAGGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6782  
2241 SerGlnGlyValIleValIleSerValIleIleTyrArgThrIleuValaglyLeuLeuPro 2260  
6783 AGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6842  
2261 HisAsnTyrAspProArgLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
6843 CATTAACCTAAGACCTTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6902  
2281 ProValIleSerIleSerValHisAspAspGluGluLeuLeuProArgValIleuAspLys 2300  
6903 CCGGTGTAGAGCATCAGCGTCCATGATGATGAGAGCTTGTGCGCGGCGCTGGAGCAA 6962  
2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleArgProIleCysVal 2320  
6963 CCGGTACAGGTGAGTTCGCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7022  
2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyIleTyrSerValArgGlyCysVal 2340  
7023 TTGTGAGACCATTCATCATCTGTGTCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7082  
2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
7083 GTGGTCTTCGCGAGATGAGAGGCAAGTCAAGTGCAGAGGCAACCAATGAGAGAGCTTGCT 7142  
2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuIleThrLeuThr 2380  
7143 GTGCTATGAGCATCTTCTCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7202  
2381 TyrValIleLeuGlyValThrLeuAlaIleLeuLeuThrPhePheLeuThrLeu 2400  
7203 TAGGTGCTTAAAGTGTACCTTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7262  
2401 LeuArgIleLeuArgSerAsnGlnIleGlyIleArgArgAsnLeuThrAlaIleuGly 2420  
7263 TTGCGATCTCTGCGCTCAACCAAGGAGCATCGAGTAACTGACAGTGCCTGGGCG 7322  
2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
7323 CTGGCTCAGGTGCTTCTCTCGGAGATCAACAGGCTGACCTCCCTTGTGCTGACA 7382  
2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuGln 2460  
7383 GTCAATGCCATCTCGCTGCACTTCTGTACTGTGACCTTTCCTGTGGCTGTGGAG 7442  
2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
7443 GCGTTGCACTGTACCGGAGCACTCAAGAGTGCAGATGTAACACCGGCGCCCAAGCGG 7502  
2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
7503 TTCTACTACATGCTGGGCTGGGCGGCGTGCCTTCATCAAGAGGCTAGCGGCGGCTG 7562  
2501 AspProGluGlyTyrGlyAsnProAspPheCysTyrLeuSerIleTyrAspThrLeuIle 2520  
7563 GACCCGAGGGGCTACGAGGAGCCGACTTCTGCGGTCTTCATCTATGACAGGCTCATC 7622  
2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
7623 TGGAGTTTCTGTGCGCGGTGCGCTTGTGCGTCTCGATGAGTGTCTTCTGATCATCTG 7682  
2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnIlePheGlyLysLysGlyProValSer 2560

7683 GCGGCCCGGCGCTCTGTGCTGCCAGCGGAGGCGCTTGTGAGAGAGAGAGTCTGCTCG 7742  
2561 GilyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrThrLeuLeuAlaLeu 2580  
7743 GGCCTGCAAGCTCTCTTGGCGGTCTCTGTGCTGAGAGCGGCACTGTGCTGTGCACTG 7802  
2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrIleuPheAlaThrCysAsnCysIle 2600  
7803 CTTCTGTCAACAGAGACACCTCTCTTCCACTACCTCTTGTGCTGAGATGAGATC 7862  
2601 GlnGlyProPheIlePheLeuSerTyrValIleuSerLysGluValArgLysAlaLeu 2620  
7863 CAGGCGCCCTTCATCTTCTCTCTATGTGTGCTTACAGAGAGTCCGAGAACACTGC 7922  
2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
7923 AGCTTGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7982  
2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
7983 ACCTGTCTTCAACTGCGCCAGCGCTTACGAGATGGCGGCTGTACAGGCTTACGGA 8042  
2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
8043 GACTGGCGCGCTCTGTGACAGCACAGTGTGCTGCGGAGAGAGTCAAGCCAGCTTACATC 8102  
2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
8103 CCTTCTTGTGAGGAGAGAGAGTCCGACTGAACTTGGCCAGAGGCGCCCTTGGCTGGGG 8162  
2701 AspProGlySerLeuPheLeuGluGlyValIleAspGlnGlnHisAspProAspThrAspSer 2720  
8163 GATCCAGGAGGCTGTCTTCTGAGAGTCAAGACAGAGCATGATCTGACAGGAGCTCC 8222  
2721 AspSerAspLeuSerLeuGluAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
8223 GACAGTGAACCTGTCTTGAAGAGCAGCAGAGTGTCTCTTATGCTTACCACTCATCA 8282  
2741 AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
8283 GACAGTGAAG 8342  
2761 TrpAspSerLeuLeuGlyProGlyValIleGluLeuProLeuHisSerThrProLysAsp 2780  
8343 TGGATAGGCTGTGCTGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8402  
2781 GlyGlyProGlyProGlyLysAlaProThrProGlyAspPheGlyThrThrAlaLysGlu 2800  
8403 GGGGCGCCAGGCGCTGGAGAGGCGCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 8462  
2801 SerSerGlyAsnGlyValAspGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
8463 AGTATGGGAGAGAGGCGCCCTTGAAGAGGCGCTGCGGAGAGAGAGAGAGAGAGAGAG 8522  
2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyIleLeuLys 2840  
8523 GAGGCGCTCTTGGGCGCCCTTCAAGGCTTCTGTGCGAGGCTCAACAAAGGCACTTAA 8582  
2841 LysLysCysLeuProThrIleSerGlyLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
8583 AAGAGGTGTCTGCCCACTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642  
2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyLysProProPro 2880  
8643 TGCAAGGGGTCTTCCCGGGGCTCTCCGCTATGAGGGGAGCGGCGGCGGCGCCCTGCC 8702  
2881 ArgProProProArgGlnSerLeuGlnGluLeuAsnGlyValMetProIleAlaMet 2900  
8703 CCGCCACGCGCCCGGAGAGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8762  
2901 SerIleLysAlaGlyThrValAspGluAspSerSerGlySerGluPheLeuPheAsn 2920  
8763 AGCATCAAGGAGGAGAGGAG 8822

OY 2921 PheleuHs 2923  
 DB 8823 TTCTGAT 8831  
 RESULT 3  
 ABX34546  
 ID ABX34546 standard; cDNA, 11762 BP.  
 AC ABX34546;  
 XX 13-FEB-2003 (first entry)  
 DT Human mdt cDNA SEQ ID 107.  
 DE  
 XX MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianemic; antiproliferic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;  
 KW psoriasis; hepatitis; gene; ss.  
 XX Homo sapiens.  
 OS  
 MO200279449-A2.  
 XX 10-OCT-2002.  
 PD 27-MAR-2002; 2002WO-US09944.  
 PE  
 XX 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 29-MAR-2001; 2001US-280068P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 17-MAY-2001; 2001US-291849P.  
 PR 19-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 P1 Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 P1 Dutour GE, Hillman JL, Yu JY, Tuason O, Yap PB, Amshy SR;  
 P1 Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klefeld Y, Gerstein EH;  
 P1 Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 P1 Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 DR WPI; 2003-058431/05.  
 DR P-PSDB; ABU11556.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis  
 PT  
 PS Claim 1; SEQ ID NO 107; 3399p + Sequence Listing; English.  
 XX  
 CC This invention describes a novel disease detection and treatment molecule  
 CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
 CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
 CC antianemic, antiproliferic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
 CC hepatitis. ABX34440-ABX34835 encode the MDDT polypeptides represented in

CC ABU11450-ABU11845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 11762 BP; 2286 A; 3738 C; 3283 G; 2455 T; 0 other;  
 Alignment Scores:  
 Pred. NO.: 0 Length: 11762  
 Score: 2894.00 Matches: 2894  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.01% Indels: 0  
 DB: 25 Gaps: 0  
 US-09-916-849a-3 (1-2923) x ABX34546 (1-11762)  
 OY 30 LeuGIYAspGlnValGlyProCyAsrSerLeuGIYSerArgIYValGlySerSerGIY 49  
 DB 2 TTGGGAGACCAAGTGGGGCCCTGTCCTTGGGGTCCAGGGGACGAGGCTTCGGGG 61  
 OY 50 AlaCyAlaProMetGIYTPleuCyProSerSerAlaSerMetPleuTYThr 69  
 DB 62 GCGTGGCCCCCAGTGGGCTGCTGTCATCTCAAGCTCAAGCTGCTGCTTACACC 121  
 OY 70 SerArgCyAsrAspAlaGIYThrGluLeuThrGIYHisLeuValProHisAspGIY 89  
 DB 122 AGCGGCTGAGGAGGATGGGGCACTGAGCTGACCTGGCACTGACCAAGATGCG 181  
 OY 90 LeuArgValITProCyProGluSerGIYAlaHisITLeProLeuProAlaProGluGIY 109  
 DB 182 CTGAGGGTGTGTGTCCAGAAATCCAGGGCCCATATTCCTCCACCAAGCTCTGAAAGGC 241  
 OY 110 CyAsrProTserCyAsrArgLeuGIYLeGIYHisLeuSerProGluGIYValLeu 129  
 DB 242 TGCCCTGGAGCTGTGCTCTCTGGGATGAGGCCACTTCTCCACAGGGCAAGCTC 301  
 OY 130 ThrLeuProGluGIYHisProCyAlaLeuValProArgLeuArgCyGlnSerCyAla 149  
 DB 302 ACACTGCCGAGGAGGACCCGTGCTTAAGGCTCCAGGCTCAAGATCCAGTCTGCAAG 361  
 OY 150 LeuAlaGlnAlaProGluLeuArgAlaGIYValArgSerProGluGIYSerLeuGIY 169  
 DB 362 CTGGCACAGGCCCCCGGGCTCAAGGGCAAGGATCAACAGAGATGCTCTGGGGGG 421  
 OY 170 ArgArgLeuArgAsnValaThrThrAlaProGluPheGluProProSerTYrGlnAlaThr 189  
 DB 422 CGTGGAAAGAAATGTAATATACAGCCCCAGTTCAGGCCCCCGAGCTACAGGCCACA 481  
 OY 190 ValProGluAsnGlnProAlaGIYThrProValAlaSerLeuArgAlaITLeaProAsp 209  
 DB 482 GTGCCGGAAGAACAGGACAGGACAGCCCGTGTGATCCCTGAGGGCCATCGACCGGAC 541  
 OY 210 GlnGlyGlnAlaGlyArgLeuGIYTYrMetAspAlaLeuPheAsrArgSerAsn 229  
 DB 542 GAGGTGAGGACAGTGTGACAGTACAGCCCATCATGAGTCCCTTGTGATACCCCTCCAC 601  
 OY 230 GlnPheSerLeuAspProValIThrGlyAlaValIThrAlaGluGluLeuAspArg 249  
 DB 602 CAGTTCCTTCCTCGAACCCAGTCACTGTCAGTAACACACAGCCGAGAGCTGAGATCGT 661  
 OY 250 GlnIThrLeuSerThrHisValPheArgValIThrAlaGlnAspHisGlyMetProArgArg 269  
 DB 662 GAGACCAAGGACCAACCGTCTTCAAGGTCAAGGCCACAGGACCAAGCCAGGCCCGACA 721  
 OY 270 SerAlaLeuAlaThrLeuThrITLeuValIThrAspThrAsnAspHisAspProValPhe 289  
 DB 722 AGTGCCTGTGCTACACTCAACATCTGTGTTACGACCAACATACATGACCTGTGTTTC 781  
 OY 290 GlnGlnGlnGluTYrITyLeuGluSerLeuArgValAsnLeuGluValGlyTYrGluValLeu 309  
 DB 782 GAGCAGCAGAGATCAAGAGAGAGCTCAAGGAGAACTTGAGGTGGCTATGAGGTGCTC 841



QY 310 ThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeu 329  
DB 842 ACTGTAGAGGCGCCAGGATGGTATGCCCTCCCAATGCCAAATATCTGATACGGCTGCTG 901  
QY 330 GluGlySerGlyGlySerProSerGlyValPheGluIleAspProArgSerGlyValIle 349  
DB 902 GAGGGGTCTGGGGGAGCCCTCTGAAAGCTTGTGAATGCAACCTGCTCTGGGGTGAATC 961  
QY 350 ArgThrArgGlyProValAspArgGluGluValGlySerTyrGlnLeuThrValGluAla 369  
DB 962 CGAACCCGTGGCCCTGTGATCGGGAAGAGTGAATCTTACCAAGCTGACGGTAGAGCA 1021  
QY 370 SerAspGlnGlyArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerVal 389  
DB 1022 AGTGACACAGGGTGGGACCCGGGTCTCGGAGTACACAGCCGCTGTTTCTTCTGCTG 1081  
QY 390 GluAspAspAsnAspAsnAlaProGlnPheSerGlyLeuArgTyrValValGlnValArg 409  
DB 1082 GAGGATGACAAATGATATGCCCCCAGTTTATGAGAGAGCGCTATGTGTCCAGGTGAG 1141  
QY 410 GluAspValThrProGlyValAlaProValLeuArgValThrAlaSerAspArgAspGly 429  
DB 1142 GAGGATGATCTCCAGGGGCCCAAGTACTCCAGTCAAGCTCGGATCGAGACAAAGGG 1201  
QY 430 SerAsnAlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeu 449  
DB 1202 AGCAATGCCGTGGTGCATATAGCATCATGATGGCAATGTCTGGGGCAGATTATCTG 1261  
QY 450 AspAlaGlnThrGlyValAlaLeuAspValValSerProLeuAspTyrGlnThrTyrLeu 469  
DB 1262 GATGCCCAATCGAGACTCTGATGTGGAGCCCTTGACTATGACGACCAAGGAG 1321  
QY 470 TyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGly 489  
DB 1322 TACACCTTACGGGTGGAGACACAGATGTGGCCGCCCTCCCACTCTTATGTCTTGGC 1381  
QY 490 LeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrPro 509  
DB 1382 TTGGTGACAGTACAGGTCTCTGATATCAACGAAATGCCCATCTTCTCTGACACCCCT 1441  
QY 510 PheGlnAlaThrValLeuGlnSerValProLeuGlyTyrLeuValLeuHisValGlnAla 529  
DB 1442 TTCCAGGCTACTGTCTCTGAGAGCGTCCCTTATGGCTACTGTGTCTCATGTCCAGGCT 1501  
QY 530 IleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTyrArgLeuAlaGlyValGlyHis 549  
DB 1502 ATCGACGCTGATGCTGTGACAAATGCCCTGGAAATACCGCTGTCTGGGGTGGACAT 1561  
QY 550 AspPheProPheThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAsp 569  
DB 1562 GACTTCCCTTCAACCATCAACATGGCACAGGCTGATCTGTGTGCTGTGAACCTGGAC 1621  
QY 570 ArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAla 589  
DB 1622 CGGAGAGAGTGAATTTCTACAGCTTTGGGGTAGAAGCTCGAGACATAGCCATCCAGCA 1681  
QY 590 LeuThrAlaSerAlaSerValSerValThrValLeuAspValAsnAspAsnProThr 609  
DB 1682 CTCACGTGCTCGGCGCAGGTGACGCGTACTGTCTGGATGTCAAGACAAATCCAAAC 1741  
QY 610 PheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerVal 629  
DB 1742 TTATCCCAACAGAGTACACAGTGGCGCTCAATAGAGATGAGCTGTGGCCACAGCGG 1801  
QY 630 ValThrValSerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSer 649  
DB 1802 GTGACGGTGTCAAGCTGTGACCGGTATGTCTCATAGTGTATCATCTTACCAAGTACCA 1861  
QY 650 GlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeu 669  
DB 1862 GGCAATACTCGAAACCGCTTCTTCATCACACGCAAAAGTGTGGGTGTATATCCCTT 1921  
QY 670 AlaLeuProLeuAspTyrTyrIleLeuGluArgGlnTyrValIleValAlaValThrAlaSerAsp 689

DB 1922 GCCCTGCCACTGACCTACAAACTTGAGCGGAGTATGTGTGGCTGTATACGGCTCCGAT 1981  
QY 690 GlyThrArgGluAspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArg 709  
DB 1982 GGCATCTGGCAGGACACGGGCAACAGTGTGGTGAATGTCACGACGCAACCCCATGT 2041  
QY 710 ProValPheGlnSerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThr 729  
DB 2042 CCTGTCTTCAAGAGTCCCACTATACGTGAATGTATATGAGACCGGCCGCGACGAC 2101  
QY 730 ThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyr 749  
DB 2102 ACGGTGTGTGATACAGCGCCACGATGAGACACAGGTGAGAAATGCCCGATCACTAC 2161  
QY 750 PheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyValValThrThr 769  
DB 2162 TTCATGAGAGACAGATATCCCAAGTTCGATCCGATCATGACAGACGAGGGGCTGTCAAC 2221  
QY 770 GlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAsp 789  
DB 2222 CAGGCTGAGCTGAGCTACGAGACCAAGTGTCTTACACCTCGCATTAAGTCTCGGAC 2281  
QY 790 AsnGlyIleProGlnLeuSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsn 809  
DB 2282 AATGCGATTCGCCAGAGTCCGACACCACTTACCTGAGATCTGTGTGAAACGAGTGAAT 2341  
QY 810 AspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValPro 829  
DB 2342 GACATATGCCCTCAAGTCTCTCGAGACTCTTACCAAGGCAAGTGTCTATGAGAGTGTCCA 2401  
QY 830 ProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArg 849  
DB 2402 CCTTCACTACCGTCTCGACATCTCAAGCCACTATGTGATTTCTGACCTTAATGGCAGG 2461  
QY 850 ValPheTyrThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThr 869  
DB 2462 GCTTCTACACCTTCCAAAGGAGGAGACATGAGACAGGTGACTTATTTGTGTGATCTCAG 2521  
QY 870 SerGlyIleValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeu 889  
DB 2522 TCAGGATCTGTGCACAGCTACGAGGCTGATGAGAGAACGTGGCCAGTATGTCTTG 2581  
QY 890 ArgAlaTyrAlaValAspLeuGlyMetProProAlaArgThrProMetGluValThrVal 909  
DB 2582 CGGCGATATGACAGTGAACAGGGGATGCCCAACCCGCACTATATGAGATGACAGTCTC 2641  
QY 910 ThrValLeuAspValAsnAspAsnProProValPheGluGlnAspGluPheAspValPhe 929  
DB 2642 ACTGTGTGGATGTGAATGACAAATCCCTGTCTTTGAGCAGAGATGATGTGATGTTT 2701  
QY 930 ValGluGluAsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAsp 949  
DB 2702 GTGAGAGAGACACCCCATTTGGGCTAGCCGTGGCCGGGTCAAGCCATCCAGAT 2761  
QY 950 GluGlyThrAsnAlaGlnIleMetTyrGlnIleValGluGlyAsnIleProGluValPhe 969  
DB 2762 GAAAGCACCAATGCCAGATATATGACAGATGTGAGGGCAACATCCCTGAGGTCTTC 2821  
QY 970 GlnLeuAspIlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArg 989  
DB 2822 CAGCTGACATCTTCTCCGGGAGACTACAGCCCTGTGTAGCTTATGACTACGAGGACCGG 2881  
QY 990 ProGluTyrValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrVal 1009  
DB 2882 CCTGAGTACGTCTCGTGTATCTCAAGCCACGTCACTCTCTGTGTGTGAGCGGGGTCAAGTCTC 2941  
QY 1010 HisValArgLeuLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeu 1029  
DB 2942 CACGTCCGCTCTCTGACCGCAATGACAAACCAATGCTGGGCAACTTTGAGATCTT 3001  
QY 1030 PheAsnAsnTyrValThrAsnArgSerSerPheProGlyGlyAlaIleGlyArgVal 1049



Dh 3002 TTCAACAACTATGTACCAATGCTCAAGACAGCTTCCTGGGGGTGCCATTGGCCGAGTA 3061  
Qy ProAlaHisAspProAsp11LeSerAspSerLeuThrTySerPheGluArgGlyValAsnGlu 1069  
Db CCTGCCATTGACCCGATATCTCAGATAGTCTGACTTACAGCTTTGACGGGGAAATGAA 3121  
Qy LeuSerLeuValLeuLeuAsnAlaSerThrGlyGluLeuValLeuSerArgAlaLeuAsp 1089  
Db CTACACCTGCTCTCTCTCAATCCCTCCACGGGTGAGCTGAGCTAAGCCGCCACTGAGC 3181  
Qy AsnAsnArgProLeuGluAlaAlaIleMetSerValLeuValSerAspGlyValHisSerVal 1109  
Db AACAAACCGGCTCTGAGGCAATGAGCGTGTGTGTGACAGCGGCTGACAGCGGTG 3241  
Qy ThrAlaGlnCysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIle 1129  
Db ACCGCCACGTGGCGCTGGTGTGACATCATCACGATGAGATCTCTACCCACAGCATC 3301  
Qy ThrLeuArgLeuGluAspMetSerProGluIleArgPheLeuSerProLeuLeuGlyLeuPhe 1149  
Db ACGCTGCCTGTGAGACATGTACCCGAGCGCTTCTGTCCACACTGAGGCTCTTTC 3361  
Qy IleGlnAlaValAlaAlaThrLeuAlaThrProProAspHisValValPheAsnAla 1169  
Db ATCCAGCGGTGGCGCCACGCTGGCCACGCCACCGAACCAGTGTGTGCTTCAACGTA 3421  
Qy GlnArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGln 1189  
Db CAGCGGACACCGACGCCCCCGGGGGCCACATCCTCAACGTAGCTGTGGTGGCCAG 3481  
Qy ProProGlyProGlyGlyGlyProProPheLeuProSerGluAspLeuGlnIleArgLeu 1209  
Db CCGCCAGGCGCCGGGGGGGGCGGCGGCTTCCTGCTCGAGGACCTTCAGAGAGGCTTA 3541  
Qy TyrlLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsp 1229  
Db TACCTCAACCGGACGCTGTGACGGCCATCTGGGACAGCGGTGTGCTGCTTCAACAC 3601  
Qy AsnIleCysLeuArgGluProCysGlnAsnIleMetArgCysValSerValLeuArgPhe 1249  
Db AACATCTGCTGGGGAGCCCTGCGAAGATACATGCGGTGTGTGTGTGTGTGCGCTTC 3661  
Qy AspSerSerAlaProPheIleAlaSerSerSerValIlePheArgProIleHisProVal 1269  
Db GACTCTCCGGGCGCTTATGCTCCTCCTCCGCTTCCTCCGACCATCACCCCGCTC 3721  
Qy GlyGlyLeuArgCysArgCysProProGlyPheThrGlyAspTyArgGluThrGluVal 1289  
Db GAGAGGCTGCGCTGCGCTGCGGCGGCTTCACGGGTGACTGCTGCGAGACCGAGGTG 3781  
Qy AspLeuCysTySerArgProCysGlyProHisGlyArgCysArgSerArgGluGly 1309  
Db GACCTCTGCTACTGCGGGCCCTGTGGCCCTCCACGGGCGCTGCGGACCGGAGGGCGG 3841  
Qy TyrThrCysLeuCysArgAspGlyTyrThrGlyGluHisCysGluValSerAlaArgSer 1329  
Db TACACCTGCTCTGT 3901  
Qy GlyArgCysThrProGlyValCysValAsnGlyGlyTyrCysValAsnLeuLeuValGly 1349  
Db GGGCTTGTGACCCCGGGGT 3961  
Qy GlyPheIleCysAspCysProSerGlyAspPheGluValProTyArgGlnValThrThr 1369  
Db GGTTCAGAGTGAGATTGCGCATCTGTGAGATCGGAAAGCCCTTATGTGCAAGGTGACAG 4021  
Qy ArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPhe 1389  
Db CGACACTTCCCGCCACTCTCTTATCACTTCCCGGCGTGGCGCACGTTTCCACTTC 4081  
Qy ThrLeuAlaLeuSerPheAlaThrArgGluArgAspGlyLeuLeuLeuTyrAsnGlyArg 1409  
Db ACCCTGGCCCTCTCTGT 4141

Qy PheAsnGluValHisAspPheValAlaLeuGluValIleGlnGlnGlnValGlnLeuThr 1410  
Db TTCAATAGAGAGATGACTTTGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 4201  
Qy PheSerAlaGlyGlnSerThrThrThrValSerProPheValProGlyGlyValSerAsp 1449  
Db TTCTCTGAGGGGAGTACACCAACGAGTGTCCCATTTCCGTCGGGAGAGTCAAGTAT 4261  
Qy GlyGlnIleThrPheThrValGlnLeuLeuTyrTyrAsnIleProLeuLeuGlyGlnThrGly 1469  
Db GGCCAGTGGCATAGGTGACGCTGAATCTGACGGGCGCTGTATGACCACTGTGGGTCAAG 4321  
Qy LeuProGlnGlyProSerGlnGlnValAlaValValThrValAspGlyCysAspThr 1489  
Db CTCCACAGGGGCCATCAGAGCAGAGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4381  
Qy GlyValAlaLeuArgPheGlySerValLeuGlyValAsnTySerCysAlaIleGlnGlyThr 1509  
Db GAGTGTGCTTGTGCTTGT 4441  
Qy GlnGlyGlySerIleValSerLeuAspLeuThrGlyProLeuLeuGlnGlyValPro 1529  
Db CAGGATGCGAGAAAGATCTGTGATCTGACGGGCGCTGTATGACGAGGGGCTGCT 4501  
Qy AspLeuProGluSerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeu 1549  
Db GACTGTCCAGAGACTTCCAGTCCGAAATCGGAGATTCGTGGCTGCATGCGAAGCTG 4561  
Qy GlnValAspSerArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrValPro 1569  
Db CAGGTGACAGCCCGGCATATGACATGCTGACTTATGCAACATAGGACCGTGCCT 4621  
Qy GlyCysProAlaValSerValAsnValCysAspSerAsnThrCysHisAsnGlyGlyThrCys 1589  
Db GGTGTCCCTGCAAGAAAGATGTGTGACAGCAACTTGCACAAATGGGGGCACTTGC 4681  
Qy ValAsnGlnIleProAspAlaPheSerCysGluCysProLeuGlyPheGlyGlyTySerCys 1609  
Db GTGAACCAAGTGGACGCGTTCACTGACGTGAGTGCCCTTGGGGGCAAGGCTGC 4741  
Qy AlaGlnGlnMetAlaAsnProGlnHisPheLeuGlySerSerLeuValAlaThrHisGly 1629  
Db GCCCAGAAATGGCCAAATCCACAGCACTTCTTGGGCGAGCCCTGTGCAATGCG 4801  
Qy LeuSerLeuProIleSerGlnProTyrThrLeuSerLeuMetPheArgThrArgGlnAla 1649  
Db CTCTGCTGCTCCATCTCCAACTGTGTACTGACCTCATGTTCGACGCGCAGGCTC 4861  
Qy AspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArg 1669  
Db GAGGTGTCCGTGCGAGGCATCACACGAGGGGGGAGACACATCATCACTGACACTGCA 4921  
Qy GlnGlyHisValMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeu 1689  
Db GAGGGCCACGTGAAGCTGAGCGTGAAGGGGACAGGGCTTCAAGGCTCTCTCTCCGCTG 4981  
Qy GlnProGlyIleArgAlaAsnAspGlyAspThrHisAlaGlnLeuAlaLeuGlyValAsn 1709  
Db GAGCCAGGCGGGCCAAATGACGTGACTGTGACCATCATCAAGCTTGGGACCTGGAGCC 5041  
Qy GlyGlyProGlyHisAlaIleLeuSerPheAspTyArgGlnGlnArgAlaGlnGlyValAsn 1729  
Db GGGGGGCTGGCCATGCTCATTTGTCTTGTGATTTATGGGAGGAGAGAGAGGGGCAAC 5101  
Qy LeuGlyProArgGluHisGlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGly 1749  
Db CTGGGCCCCCGGCTGACATGTGTGCACTTGAACATCAAGTGGGCGGAATACCTGGG 5161  
Qy ProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlnValArgValSerAsp 1769  
Db CAGCCGCGGTGTGGCCCTTGTGGGCTTGTGGGCTTGTGAGGGTGTGCGGGTGAAGCAT 5221

QY 1770 ThrProGluGlyValIasnSerLeuAspProSerHisGlyIleSerIleasnValGluGln 1789  
DB 5222 AGCCCAAGAGGGGGTTAAAGCCCTGGATCCAGCCAAAGGGAGAGCATCAACGGAGGCAAA 5281  
QY 1790 GlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSer 1809  
DB 5282 GGGCTGAGCTGGCTTACACCTGTGACTCAACCCGTCGCTCAACAGCTATTTGCAAC 5341  
QY 1810 AsnAspTrpAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAsnAsnCysThr 1829  
DB 5342 AACGACTGGGAGACGATATCTCTGCACTGTGATCCAGGTTACATATGTGACAACTGACT 5401  
QY 1830 AsnValCysAspLeuAsnProCysGluHisGlnSerValCysThrArgIleProSerAla 1849  
DB 5402 AATGTGTGTGACCTGTAACCCGTGTGAGCAACAGTGTGTGTACCCGCAAGCCCAAGTGGC 5461  
QY 1850 ProHisGlyTyrThrCysGluCysProProAsnTyrIleuGlyProTyrCysGluThrArg 1869  
DB 5462 CCCCATGGCTATACCTGCAAGTGTCCCCCAATTACCTTGGGCACTATCTGTAGAGCAAG 5521  
QY 1870 IleAspGlnProCysAspProArgGlyTyrTrpGlyHisProThrCysGlyProCysAsnCys 1889  
DB 5522 ATTGACACAGCTTGTCCCGGTGGCTGTGGGAGCATCCCAATGTGGCCCATGCACTGT 5581  
QY 1890 AspValSerIleArgIlePheAspProAspCysAsnIleThrSerGlyIleCysHisGlyVal 1909  
DB 5582 GATGTGAGCAAAAGCTTTGACCACTGCACTGCACAAAGACGCGGCAAGTGGCCACTGCAAG 5641  
QY 1910 GluAsnHisGlyArgProProGlySerProThrCysIleuLeuCysAspCysTyrTrpProThr 1929  
DB 5642 GAGAACCACTACCGGCCCCAGGCAAGCCCCACTGCTCTTGTGTGATCTGTACCCCA 5701  
QY 1930 GlySerLeuSerArgValCysAspProGluAspGlyGlnCysProCysIleAspProGlyVal 1949  
DB 5702 GGGCTCTTGTCCAGAGTCTGTGACCTGAGGAGTGGCAAGTGTCAAGCAAGCAAGGTGTC 5761  
QY 1950 IleGlyArgGlnCysAspArgCysAspAsnProPheAlaGluValThrThrAsnGlyCys 1969  
DB 5762 ATCGGGCGTCAAGTGTACCGGTGTGACACCTTTGTGTGAGGTGCAACCAATGGCTGT 5821  
QY 1970 GluValAsnTyrAspSerCysProArgAlaIleGluAlaGlyIleTrpTrpProArgThr 1989  
DB 5822 GAAAGTAATTATGACAGCTGCCCCAGAGCATTTGAGGCTGGATCTGTGGCCCCCTTACC 5881  
QY 1990 ArgPheGlyLeuProAlaAlaAlaProCysProGlyGlySerPheGlyThrAlaValArg 2009  
DB 5882 CGCTTCGGGCTGCTGCTGCTGCTGCTGCTGCCAAAGGCTCTTTGGAGCTGTGGCCG 5941  
QY 2010 HisCysAspGlnHisArgGlyTyrLeuProProAsnLeuPheAsnCysThrSerIleThr 2029  
DB 5942 CACTGTGATGAGCAAGGGGGTGGTCTCCCCCAAACTTCAACTGCAAGTCCATCCACC 6001  
QY 2030 PheSerGlyLeuGlySerGlyPheAlaGluArgLeuGlnArgAsnGlySerGlyLeuAspSer 2049  
DB 6002 TTCTCAAGACTGAAGGGCTTCCGTGAGCGGCTTACAGCGAAATGAGTCAAGGCCATGACTCA 6061  
QY 2050 GlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisGThrAlaGlyTyr 2069  
DB 6062 GGGCGCTCCCAAGCAAGTACGCTGCTGCTGCGCAAGGCCACGACGACACAGCTGGCTAC 6121  
QY 2070 PheGlySerAspValIleValValAlaTyrGlnIleuAlaThrArgLeuLeuAlaHisGlnSer 2089  
DB 6122 TTGGGAGGACAGTCAAGGTGGCTTACAGCTGGCCCAAGGGGTGGCTGGCCCAAGAGG 6181  
QY 2090 ThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlyIleuAsnLeu 2109  
DB 6182 ACCCAAGGGGGCTTTGGGCTGTCTGCAACAAGAGTGTGACTTCACTGAATCTGCTGG 6241  
QY 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnIleArgHisGThrGlyIleuIleGlnGln 2129  
DB 6242 CGGGTGGGACGCGCTCTGCAACAGCCAAAGCGGCACTGGAGCTGAATCCAGCAG 6301  
QY 2130 ThrGlnGlyIleThrAlaThrLeuLeuGlnHisGThrGlyIleuAlaThrAlaSerAlaLeuAla 2149

DB 6302 ACAGAGGGTGGACCGCTGGCTGTCCAGCACTATGAGGGCTTACGCAAGTGCCTGGCC 6361  
QY 2150 GlnAsnMetArgHisGThrTyrLeuSerProPheThrIleValThrProAsnIleValIle 2169  
DB 6362 CAGAACATGGGACACCTTACCTTAAGCCCCCTTACCAATGTGTACCGCCCAACTTGTCTAC 6421  
QY 2170 SerValValArgLeuAspArgGlyAsnPheAlaGlyAlaIleValLeuProArgTyrGluAla 2189  
DB 6422 TCCGTAGTGGCTTGGCAAAAGGAACTTTGTGGGGCCAAAGTGCCTCCGCTACAGAGGC 6481  
QY 2190 LeuArgGlyGluGlnProProAspLeuGlnThrThrValIleLeuProGluSerValPhe 2209  
DB 6482 CTGCTGGGAGAGCAAGCCCCGACCTTGAACAAAGCATTTCTGCTGAGTCTGTCTTC 6541  
QY 2210 ArgGluThrProProValValArgProAlaGlyProGlyGluAlaGlnIleProGluGlu 2229  
DB 6542 AGAGAGAGCCCCCGTGTGTACAGGCCCGGAGAGGCCCGAGAGGCCAAGAGAGAG 6601  
QY 2230 LeuAlaArgArgIleArgArgHisPProGluLeuSerGlnGlyAlaValAlaSerVal 2249  
DB 6602 CTGGCAAGGCAACAGCAAGGCAAGGCAAGGCTGAGCCAGGGTGAAGCTGTGGCCAGCGTC 6661  
QY 2250 IleIleTyrArgThrIleuAlaGlyLeuLeuProHisAsnTyrAspProAspIleArgSer 2269  
DB 6662 ATCATCTACCGCACCTGGCGGGCTACTGCTCTCAATCACTGACCTGACAAAGCGCAGC 6721  
QY 2270 LeuArgValProIleArgProIleIleAsnThrProValIleSerValHisAsp 2289  
DB 6722 TTGAGAGTCCCAAAAGCCCAATCATCAACACCCGTGTGAGATGACGCTCCATGAT 6781  
QY 2290 AspGluGluLeuLeuProArgAlaLeuAspIleProValThrValGlnPheArgLeuLeu 2309  
DB 6782 GATGAGAGGCTTGTCCCGGGCCCTGGACAAACCCGTGACAGTTCAGCTGCTGCTG 6841  
QY 2310 GluThrGluGluArgThrIleCysValIlePheThrAsnHisSerIleuValSer 2329  
DB 6842 GAGACAGAGAGCGGACCAAGCCCATCTGATCAACACCCGTGTGAGATGACGCTCCATGAT 6901  
QY 2330 GlyThrGlyGlyTyrPheSerAlaArgGlyCysGluValIlePheArgAsnGlnSerHisVal 2349  
DB 6902 GGCACAGGTGGCTGTGGCCAGAGGCTGTGAAGTGTCTTCCGCAATGAGAGCCACGTC 6961  
QY 2350 SerCysGlnCysAsnHisMetThrSerPheAlaValIleMetAspValSerArgArgIle 2369  
DB 6962 AGCTCCAGTGTCAACACATGACAGCTGCTGCTGTGATGAGAGTTCCTCGGGGAG 7021  
QY 2370 AsnGlyGluIleLeuProLeuIleThrIleuThrTyrValAlaIleuGlyValThrIleuAla 2389  
DB 7022 AATGGGAGATCTGCACTGAAGACCTGAATACGTAGCTTGAAGTGTCACTTGGCT 7081  
QY 2390 AlaLeuLeuLeuThrPhePheLeuThrIleuLeuArgIleLeuArgSerAsnGlnHis 2409  
DB 7082 GCCCTTGTGCTCACTTCTTCTCTCACTGCTTGGATCTGCGCTCCCAACCAAC 7141  
QY 2410 GlyIleArgArgAsnLeuThrAlaAlaIleuGlyLeuAlaGlnIleuValPheLeuLeuGly 2429  
DB 7142 GGCATCCCAAGTAACTAAGCTGAGCTGCGGCTGCTGAGCTGTCTCTCTGAG 7201  
QY 2430 IleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeu 2449  
DB 7202 ATCAACAGGCTGACCTCTTGTGCTGCAAGATGATGCAATGCTGCTGCACTTCTG 7261  
QY 2450 TyrLeuCysThrPheSerTrpAlaLeuLeuGluAlaLeuHisIleuTyrArgAlaLeuThr 2469  
DB 7262 TACCTGTGACCTTTCTGGGCTGTGAGGCTTGCACCTGTACCGGGCACTCACT 7321  
QY 2470 GluValArgAspValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyVal 2489  
DB 7322 GAGGTGGGATGTAAACGAGCCCAAGGCTTCTACTAATGCTGGGCGGGCGTG 7381  
QY 2490 ProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAsp 2509

Db 7382 CCTGCTTCATACAGAGGCTGAGCCGTGGGCTGAGACCCCGAGGCTACACCCCTGAC 7441  
 Qy PheCysTrpLeuSerLeuTyrAspThrLeuIleTyrPheSerPheAlaGlyProValAlaPhe 2529  
 Db 7442 TTCGCTGCTCTCCATCATATGACAGCGCTGAGAGTTTGGCTGGCCGGTGGCTTT 7501  
 Qy AlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGln 2549  
 Db 7502 GCCGCTCGATGAGTGTCTTCTCTGATACATCCGCGGCGCCGGGCTCTGTGTCTGCCAG 7561  
 Qy ArgGlnGlyPheGlnLysGlyProValSerGlyLeuGlnProSerPheAlaValLeu 2569  
 Db 7562 CGGCGAGGCTTGGAGAAAGAGTCTGTCTGGGCTGAGCGCTCTCTGCGCGTCTTC 7621  
 Qy LeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAspSerPheThrLeuLeu 2589  
 Db 7622 CTGCTGCTGAGGCGCACGCTGGCTGGACCTGCTCTGTCTGACAGAGACCCCTCTTC 7681  
 Qy PheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyr 2609  
 Db 7682 TTCCACTACCTCTTGTCTACCTGCAATTGCATCCAGGGCCCTTCATCTCTCTCAT 7741  
 Qy ValValLeuSerLysGlnValArgValAlaLeuLysLeuAlaCysSerArgLysProSer 2629  
 Db 7742 GTGGTGTCTTACCAAGAGGTCGGAAGCACTCAAGCTTGGCTGCAAGCCGCAAGCCAGC 7801  
 Qy ProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsnCysProSerPro 2649  
 Db 7802 CCTACCTCTGCTGACCAACCAAGTCCACCTGACCTGTCTCTCACTGCCCCAGCCCC 7861  
 Qy TyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThr 2669  
 Db 7862 TACCGAGATGGGGCGGCTCTACCAAGCCCTACCGAGACTGGCGGCTCTCTGCAAGCACCC 7921  
 Qy SerArgSerGlyLysSerGlnProSerTyrIleProPheLeuLeuArgGlnLysSerAla 2689  
 Db 7922 AGTCGCTCGGCGCAAGAGTCAAGCCAGCTACATCCCTCTTCTGAGGAGAGAGTCCGCA 7981  
 Qy LeuAsnProGlyGlnGlyProProGlyLysGlnLysAspProGlySerLeuPheLeuGlnGly 2709  
 Db 7982 CTGAACCTTGGCCAGAGGCCCCCTGGCTGGGGGATCCAGGCGAGCTGTCTCTGAGAGGT 8041  
 Qy GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnLysAsp 2729  
 Db 8042 CAAGCCAGCAGCATGATCTGACACGCACTCCGACATGACCTGTCTTGAAGAGCAC 8101  
 Qy GlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnGln 2749  
 Db 8102 CAGAGTGGCTCTTATGCTCTTACCCATCATCAGACAGTGAAGAGAGAGAGAGAGAG 8161  
 Qy GlnGlnGlnLysAlaPheProGlyGlnGlnGlyTrpAspSerLeuLeuGlyProGlyVala 2769  
 Db 8162 GAAAGAGAGGCGGCTCTCTGAGAGAGGAGGAGGAGTACCTGTGGGAGCTGAGAGCA 8221  
 Qy GlnArgLeuProLeuHisSerThrProLysAspGlyGlyProGlyLysAlaPro 2789  
 Db 8222 GAGAGACTGCGCCCTGCAAGTACTCCCAAGATGGGGGCCCAAGGCGCTGGAAGGCCCC 8281  
 Qy TrpProGlyAspPheGlyThrThrAlaLysGlnLysSerGlyAsnGlyAlaProGlnGln 2809  
 Db 8282 TGGCCAGAGGACTTTGGAGCAGCAGCAAAAGAGTACTGCAACGCGGCGCCCTGAGAGAG 8341  
 Qy ArgLeuArgGlyLysAlaPheAlaLeuSerArgGlnGlyLysSerLeuGlyProLeuProGly 2829  
 Db 8342 CGGCTGCGGAGAGATGAGATGCTGTCTGAGAGGGGGTCCCTTAGGGCCCTTCACAGGC 8401  
 Qy SerSerAlaGlnProHisLysGlyIleLeuLysLysCysLeuProThrIleSerGln 2849  
 Db 8402 TCTTGTGCGCAGCCTCAAAAGGATCTTTAAGAAAGATGTCTGCCACATCAGCAGAG 8461  
 Qy LysSerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSerArgGlySerSer 2869  
 Db 8462 AAGAGCAGCTCTGCGGCTCTCCCTGAGAGCATGCAAGGAGTCTTCCCGGCGCTCTCTCC 8521

Qy 2870 AlaSerGlnGlySerArgGlyGlyProProProAspProProProArgGlnSerLeuGln 2889  
 Db 8522 GCTAGTAGAGGACCCGAGGAGCCCTCCCTCCGCGCACCGCCCGGAGAGCTTCAG 8581  
 Qy 2890 GlnGlnLeuAsnGlyValMetProIleAlaMetSerIleLysAlaGlyThrValAspGln 2909  
 Db 8582 GAGCAGCTGAACGGGCTGATGCCATGCGCATGAGATCAAGAGCAGGACGCTGATGAG 8641  
 Qy 2910 AspSerSerGlySerGlnPheLeuPhePheAsnPheLeuHis 2923  
 Db 8642 GACTGTCAGGCTCCGAATTTCTTCTTAACCTCTGCAAT 8683  
 RESULT 4  
 ID AS11677 standard; cDNA; 8871 BP.  
 AC AS11677;  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human Flamingo cDNA sequence.  
 KW Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
 KW signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
 KW obesity; hypertension; hyperextension; urinary retention; angina pectoris;  
 KW myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
 KW benign prostatic hypertrophy; psychotic disorder; neurological disorder;  
 KW malignant depression; delirium; dementia; severe mental retardation; ss;  
 KW Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 KW antifungal; antiviral; antiprotazoal; anti-HIV; anorectic; antianorectic;  
 KW antiparkinsonian; cardiac; cerebroprotective; neuroprotective;  
 KW antidepressant; anticonvulsant; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1..8771  
 FT /tag= a  
 FT /product= "Human Flamingo protein #1"  
 XX  
 XX MO200161003-A1.  
 XX PD 23-AUG-2001.  
 XX PF 19-FEB-2001; 2001M0-G800680.  
 XX PR 19-FEB-2000; 2000GB-0004196.  
 XX PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX PT Testa TT;  
 XX DR WPI: 2001-502792/55.  
 XX DR P-PSDB; AAU07053.  
 PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 PS Claim 5; Page 25-28; 66pp; English.  
 CC The sequence represents a cDNA which encodes a human Flamingo  
 CC polypeptide. Flamingo is a member of the G-protein coupled receptor  
 CC family, which is involved in signal transduction pathways. By screening  
 CC to identify compounds that stimulate or inhibit the function or level of  
 CC the protein, treatments can be developed for various diseases and  
 CC bacterial, fungal, protozoan and viral infections, including HIV, cancer,  
 CC diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease, acute  
 CC heart failure, hypotension, hypertension, urinary retention, angina  
 CC pectoris, myocardial infarction, stroke, ulcers, allergies and benign  
 CC prostatic hypertrophy. Also treatable are psychotic and neurological

CC disorders such as anxiety, schizophrenia, manic depression, delirium,  
CC dementia, severe mental retardation, Huntington's disease and Gilles de  
CC la Tourette's syndrome.

XX Sequence 8871 BP, 1712 A, 2859 C, 2587 G, 1713 T, 0 other:

## Alignment Scores:

Pred. No.:	0	Length:	8871
Score:	2837.00	Matches:	2837
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.06%	Indels:	0
DB:	22	Gaps:	0

US-09-916-849a-3 (1-2923) x AAS11677 (1-8871)

```

QY 1 MetAAGSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
DB 1 ATGCGAGGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 60
QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
DB 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 41 GlySerArgGlyArgGlySerSerGlyValGlyValProMetGlyTTrLeuCyProSer 60
DB 121 GGGTCCAGGGGAGCAGAGGCTCTTGCGGGGGCTGCGCCCATGGGCTGCTGCTCATCC 180
QY 61 SerAlaSerAspLeuThrLeuThrThrSerArgGlyAspAlaGlyThrGluLeuThr 80
DB 181 TCAGGCTCAACCTTGGCTCTTACACACACCGCTGAGGAGATGCGGCACTAGCTGACT 240
QY 81 GlyHisLeuValProHisAspGlyLeuArgValTTrCyProGluSerGluAlaHis 100
DB 241 GGCACCTGTGATCCCAACACATGAGCTGAGGGTTTGGTTCAGAAATCCAGGCCCAT 300
QY 101 TLeuProLeuProProAlaProGluGlyCyProTTrSerCyArgLeuLeuGlyTleuGly 120
DB 301 ATTCCCTTACCAACAGCTCTTGAAGGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTG 360
QY 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGluHisProCyLeuLeuVal 140
DB 361 GGCACCTTTCACCAAGGAGCAAGCTCACTGCGGAGAGACCCGCTTAAAGGCT 420
QY 141 ProArgLeuArgGlySerGlnSerCyValLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160
DB 421 CCACGGCTCAGATGCAATGCTGCAAGCTGCAAGGCCCCCGGGCTCAGGGGAGGGA 480
QY 161 ArgSerProGluGluSerLeuGlyValArgGlyValArgAspValAsnThrAlaProGln 180
DB 481 AGGTCAACAGAAAGATCCCTGGGTGGCGCTCGGAAAGAAATGTAAATACAGCCCCCAG 540
QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200
DB 541 TTCACGCCCCCAAGCTACCAAGCCACAGTGCAGGAACCAAGCAGCAGCAGCCCCCTGTT 600
QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGlyTTrMet 220
DB 601 GGATCCCTGAGGGCCATCGACCCGAGCAGAGGGTGAAGGCACTGAGATCAACCAAG 660
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
DB 661 GATGCCCTCTTATATATGCGCTCAACAGATTCCTCTCTGAGCCCAATCTCATCTGTGCA 720
QY 241 ValTTrThrAlaGluGluLeuAspArgGlyThrThrLeuSerThrHisValPheArgValThr 260
DB 721 GTTACCAACGCGAGAGAGCTGATCGTGACCAAGACCAACCACTCTTCAAGGGTCAAG 780
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280
DB 781 GCGCAGAGACCAAGGATGCGCCGACAGATGCGCTGCTACACTCACTCACTTGGTTACT 840
QY 281 AspThrAspAspHisAspProValPheGluGlnGlnGlyTTrLeuSerLeuAspGlu 300

```

```

DB 841 GACACCAATGACCAATGACCTGTGTGACAGCAGAGGTACAGAGAGAGCTCAGGAG 900
QY 301 AsnLeuGluValGlyTTrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320
DB 901 AACCTGAGGTGTGCTATGAGGTCTCATCTGACGGGCAAGAGATGATGATGATGATGATG 960
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyGlySerProSerGluValPhe 340
DB 961 AATGCCATATTTCTGTACCGCTGCTGAGAGGGGTCTGGGGGCAAGCCCTTGAACTCTT 1020
QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360
DB 1021 GAGATCGACCTCGCTGCGGGGTATCGAACCCTGGGCTCTGTGATCTCGGAAAGAGGTG 1080
QY 361 GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyAspProGlyProArgSer 380
DB 1081 GAATCTACAGCTGACCGGTAGAGGCAAGTACCAAGGTCGGAGACCCGGGTCTCTGGAGT 1140
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPheSer 400
DB 1141 ACCACGCGCGTGTCTTCTTCTGTGAGAGATGACATGATATATGCCCCCAGTTTATG 1200
QY 401 GlyValArgTyrValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
DB 1201 GAGAAAGCCTATGTGTCTCAGGTGAGGAGGATGTGACTCCAGGGGCCCCAGTACTCCGA 1260
QY 421 ValThrAlaSerAspAspAspGlySerAsnAlaValHisTyrSerIleMetSer 440
DB 1261 GTACAGCTCGATCGACAGACAGAGGAGCAATGCGTGTGCTATAGCATCATGAGT 1320
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
DB 1321 GGCATATGCTGGGGAGAGTTTATCTGATGCCAGACTGAGAGCTGAGATGTGGTGAAC 1380
QY 461 ProLeuAspTyrGluThrThrThrLeuGlyTTrThrLeuArgValArgAlaGlnAspGly 480
DB 1381 CCTCTTGTATGAGACCAACAGAGATGACCCCTGACGGGTGCAAGCAGAGATGTGTGC 1440
QY 481 ArgProProLeuSerAspValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
DB 1441 CGTCCCCACTCTTATATGTCTCTGCTGAGTGTGACAGTCTGATATCAACAG 1500
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
DB 1501 AATGCCCATCTTCTGTGACAGCCCTTCCAGGCTATGCTGTGAGAGAGGCTCCCTTA 1560
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
DB 1561 GGCTACCTGTGTTCTTCATGTCAGGCTATGACCGCTATGCTGTGACATATGCCCCCTG 1620
QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
DB 1621 GAATTCGCGCTGTGCGGGGTGGAGATGATCTTCCCTTCAACATCAACATGAGCAGAGC 1680
QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
DB 1681 TGGATCTCTGTGGCTGTGAATCGGACCGGAGAGATGATTTCTTACAGCTTGGGGTGA 1740
QY 581 GluIleAspAspHisGlyTTrThrProAlaLeuThrAlaSerAlaSerValThrVal 600
DB 1741 GAAAGCTCAGACCAATGAGCACTTCAGACTCATCTGCGGCAAGTCAAGGTGACTGTTC 1800
QY 601 LeuAspValAsnAspAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
DB 1801 CTGATGTCAACAGCAACATCAACCTTTCACCAACAGATACAGAGTGGGCTCAT 1860
QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
DB 1861 GAGATGAGAGCTGTGGGACCAAGGTGTGACGGTGTGACCGGTGATGCTCAT 1920
QY 641 SerValIleMetTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660

```

Db	1921	AGTGTCAATCACTTACCAAGATCAACAGTGGCAATACTGAAACCGCTTCTCATCAACGAC	1980
Qy	661	GLNserGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLyLeuGln	680
Db	1981	CAAGTGGTGGTGGGCTGTATCCCTTCCCTGGCCACTGAGACTACAACTTGAAGCGGAC	2040
Qy	661	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal	700
Db	2041	TATGTGTGGTGGTTCACCGCTCCGATGGACTCGGACAGACACGGCACAGTTGTGGTG	2100
Qy	701	AsnValThrAspAlaAspThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720
Db	2101	AATGTCAACGACGGCCAAACCCCATGTGCTGTCTTTCAGAACTCCCACTTAACAGTAGAT	2160
Qy	721	ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740
Db	2161	GTTATATAGAGACCGGCGCGGACAGCACCGGTGGTCTGATCAGCGCCACCGATAGAGAC	2220
Qy	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2221	ACAGGTGAGATGCGCCGATCACTTACATGAGAGACACATCCCAATTCGCGATC	2280
Qy	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAsnGlnValSer	780
Db	2281	GATGAGACACAGGGGGCTGTCAACCAAGGTGAGCTGAGACTCGAAGACCAAGTGTCT	2340
Qy	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrThrTyr	800
Db	2341	TACACCCCTGGCGCATTACTGCTCGGACCAATGGCATTCCCGAAGTCCGACACACTAC	2400
Qy	801	LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr	820
Db	2401	CTGGAGATCCTGGTGAACGACGTGAATGACATGACCCCTCAGTTCTCGGAGACTCTCAC	2460
Qy	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	2461	CAGGGCAAGTGTCTATGAGAGATGTGCACCCCTTCACTACCGCTCGCAGATCTCAGCACT	2520
Qy	841	AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyGluAspAspGly	860
Db	2521	GATGTGATTTCTGGACTTAATGCGAGGGTCTTCTACACTTCCAGAGAGCGGACGATGGA	2580
Qy	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValAlaGlnThrLeuArgArgLeuAsp	880
Db	2581	GACGGTGACTTATATGTTGAGTTCACACGTCAAGGACATCGTGCAACGCTACCGAGCTGGAT	2640
Qy	881	ArgGluAsnValAlaGlnTyrValValLeuArgAlaTyrAlaValAspGlyMetProPro	900
Db	2641	CGAAGAAACGTGGCCCAAGTATGTCTTGGGCAATATGACAGGACCAAGGGGATCCCCCA	2700
Qy	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal	920
Db	2701	GCCCGCACACTTAATGGAAGTACAGTCACTGTGTGGATGTGAATGACAAATCCCCCTGTC	2760
Qy	921	PheGlnGluAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAlaVal	940
Db	2761	TTTGAGCAGATGAGTTGATGTGTGTTGTGGAAAGAAACAACCCCAATTGGGCTAGCGGTG	2820
Qy	941	AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGlnIle	960
Db	2821	GCCCGGTGCACAGCCACGACCCCGATTAAGACCAATGCCAGATTATGTACCAATTT	2880
Qy	961	ValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePheSerGlyGlnLeuThrAla	980
Db	2881	GTGAGAGGCAACATCCCTGAGGTCTTCCAGCTGACATCTTCTCCGGGAGACTACAGCC	2940
Qy	981	LeuValAspLeuAspTyrGluAspArgProGlnTyrValLeuValIleGlnAlaThrSer	1000
Db	2941	CTGGTAGACTTACACTACAGGACCGGCTGATGACGTCCTGGTCAATCAGGCCACGCA	3000
Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	3001	GCTCCTCTGGTGAACGGGGCTTACAGTCCACGTCCGCTCTTGAACCGGAATGACAAACCA	3060

QY	1021	ProValIleuGIYAsnPhenIleuIleuPheAsnAsnTyrValThrAsnArgSerSer	1040
Db	3061	CGAGGCTGGGGCAACTTGGATGCTTTTCAACAACAATGTCAACAATCGCTCAAGACGC	3120
QY	1041	PheProGIYGIYAlaIleGIYArgValProAlaHisnProAspIleSerAspSerIleu	1060
Db	3121	TTCCCTGGGGGGTGGCATTGGCCGAGTACTGCGCATGACCTTGATATCTCAGATAGTCTG	3180
QY	1061	ThrTyrSerPheGIuArgIYAsnGIuIleuSerIleuValIleuIleuAsnAlaSerThrGIY	1080
Db	3181	ACTTACAGCTTTGAGCGGGGAAATGAATCACTGAGCTGGCTCTCATATGCTCCACGGGT	3240
QY	1081	GIuIleuIleuSerArgAlaIleuAspAsnAsnArgProIleuGIYAlaIleuSerVal	1100
Db	3241	GAGCTGAAGCTTAAGCGCGGCACTGTACAACAACCGGCTCTGGAGGGCCATCATAGCGTG	3300
QY	1101	IleuValSerAspGIYAlaHisSerValThrAlaGIuIleuValIleuValThrIleIle	1120
Db	3301	CTGGGTGACAGCGCGATACACAGGGTGAACGGCCGAGTGGCGGTGGGTGACATATC	3360
QY	1121	ThrAspGIuMetIleuThrHisSerIleThrIleuArgIleuIleuAspMetSerProGIuArg	1140
Db	3361	ACCATGAGAGCTTACCGCACCCACAGCATCACTGTGGCTGTGAGGACAATGTACCCGACGCG	3420
QY	1141	PheIleuSerProIleuIleuGIYIleuPheIleGIuAlaValAlaAlaThrIleuAlaThrPro	1160
Db	3421	TTCTGTGACCACTGCTAGAGGCTTTATCTCAAGCGGTGGCCGACGCTGACACGCCA	3480
QY	1161	ProAspHisValValIlePheAsnValGIuArgAspThrAspAlaProGIYGIYHisIle	1180
Db	3481	CCGAGACCACTGGTGGTCTTCAACGTACAGGGGACACCAACGCGCCCGGGGGCCATC	3540
QY	1181	IleuAsnValSerIleuSerValGIYGIuIleuProProGIYProGIYGIYProProPheIleu	1200
Db	3541	CTCAACGTGAACCTGTGTGGTGGGCGACCGCGACAGAGCGCGGGGGCGGCGCTTCTCG	3600
QY	1201	ProSerGIuAspIleuGIuArgIleuTyrIleuAsnArgSerIleuIleuThrAlaIleSer	1220
Db	3601	CCCTCTGAGGACCTGACGAGAGCGGCTTATCACTCAACCGGACGCGTGAACGGCATCTCG	3660
QY	1221	AlaIleuArgValIleuProPheAspAspAsnIleCysIleuArgIleuProCysIleuAsnTyr	1240
Db	3661	GCAACAGCGCGGTGCGCTTCAACAGAACATCTGCTGGGGGAGCCCTCGAACAACTAC	3720
QY	1241	MetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260
Db	3721	ATGGGCTGGGTGGGTGGTGGCTTGAATCTCTCGCGGCCCTTCATCGGCTCTCTCC	3780
QY	1261	ValIleuPheArgProIleHisProValGIYIleuArgCysArgCysProProGIYIle	1280
Db	3781	GTGCTCTTCCGGCCCATCAACCCCGGTGAGAGGCGTGGCGCTGCGCGCCGCGGCTTC	3840
QY	1281	ThrGIYAspTyrCysGIuThrGIuValAspIleuCysTyrSerArgProCysGIYProHis	1300
Db	3841	ACGGGTGATCTACGTGAGAACCGAGGTGAGACCTGTGATCTGCGGGCCCTGTGGCCCCAC	3900
QY	1301	GIYArgCysArgSerArgIleuGIYTYrThrCysIleuCysArgAspGIYTYrThrGIY	1320
Db	3901	GGGGCGTGGCGACGCGGAGGGGGGCTTAACAACCTGCTGTGTGATGGCTACACGGGT	3960
QY	1321	GIuHisCysArgIleuAlaSerAlaArgSerGIYArgCysThrProGIYValCysIleuAsnGIY	1340
Db	3961	GAGACACTGTGATGGATGTCTCGCTCAAGGCGGTGACCCCGGGTGTCTCAAGAAAGGG	4020
QY	1341	GIYThrCysValAsnIleuIleuValGIYIleuPheIleCysAspCysProSerGIYAspPhe	1360
Db	4021	GGCACCTGTGTCAACTGCTGGTGGGGGTTTCAAGTGCATTTGCCATCTGAGAGCTTC	4080
QY	1361	GIuIleuProTyrCysGIuValThrThrArgSerPheProAlaHisSerPheIleuThrPhe	1380
Db	4081	GAGAAAGCCCTAATGCGCAGTGAACACCGGAGCTTCCCGGCCCACTCTTCATCACTT	4140

QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLyseGluArg 1400  
 DB 4141 CCGGGGCTCGCGCAGGAGGTTTCCACTTCACTGCGGCTCTCTGTTGGCCAAAGAGGGCC 4200  
 QY 1401 AapGlyLeuLeuLeuLeuTyrransGlyArgPheAsnGluLysHisAspPheValAlaLeuGlu 1420  
 DB 4201 GACGGGTGCTGTGTATGACATGGCGCTTTCATGAGAACATGATCTTGTGCGCTCCGAG 4260  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLysThrThrThrValSer 1440  
 DB 4261 GTGATCCAGAGACAGGTCCAGCTCACTTCTCTGCGAGGGAGTACACCAACCGGTGTCC 4320  
 QY 1441 ProPheValProGlyGlyValSerAapGlyGlnTrpHisPheThrValGlnLeuLysTyrr 1460  
 DB 4321 CCATTGCTGCTCCGAGAGATCAAGTATGGCCAGTGGCATACGGTGGACGTGAATATCTAC 4380  
 QY 1461 AenLysProLeuLeuGlyGlnThrArgLysLeuProGlnGlyProSerGluGlnLysValAla 1480  
 DB 4381 AATTAAGCCACTGTGTGGTGAACAGGGCTCCCAAGGGGCCCATCAGAGCAGAAAGGTGGCT 4440  
 QY 1481 ValValThrValAspGlyCysAapThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB 4441 GTGGTGAACGTGGATGGCTGTGTACACAGAGTGGCTTGGCTTGGATCTGTCTCTGGCC 4500  
 QY 1501 AenTyrrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLysSerLeuAapLeuThr 1520  
 DB 4501 AACTACTCTGTGTGCTCCAGAGGCCACCCAGGGTGGCAGCAAGAAAGTCTGTGATCTGACG 4560  
 QY 1521 GlyProLeuLeuLeuGlyGlyValValProAapLeuProGlnSerPheProValArgPheLys 1540  
 DB 4561 GGGCCCCCTGCTACTAGGGCGGGGCTGTGACTGCGCCGAGAGCTTCCCACTCCGAATGGCG 4620  
 QY 1541 GlnPheValGlyCysMetArgAenLeuGlnValAspSerArgHisIleAspMetAlaAap 1560  
 DB 4621 CAGTTCGTGGGCTGCATGGCGAACTGCGAGTGGAGACCGCGGCACTAGACATGGCTGAC 4680  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysValAsnValCysAapSer 1580  
 DB 4681 TTCATTGGCCCAACATGGCACCGGTGCTGGCTCGCCGACAGAAAGAACTGTGTGACAGC 4740  
 QY 1581 AenThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAapAlaPheSerCysGluCys 1600  
 DB 4741 AACACTTGGCACAATGGGGGCACTTGCATGAAACAGTGGGAGCGCTTCAAGCTGCGAGTCC 4800  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
 DB 4801 CCCCTGGGCTTTGGGGGCAAGAGCTGCGCCGACAGAAATGGCCAAATCACAAGCACTTCCG 4860  
 QY 1621 GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrrLeu 1640  
 DB 4861 GGCAGCAGGCTGGGTGGCTGGCATGGCTCTGCTGCTCCATCTCCCAACCCCTGTGTAACCTC 4920  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAapGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 DB 4921 AGCTTATGTTCGCGACGGCCAGGCCAGAGGTGTCTGTGAGGCCATCAACAGGGGG 4980  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 DB 4981 CGAGAGCAATCACTTACAGCTACAGAGGGGCACTGATGCTGAGCCTGGAGGGCAACA 5040  
 QY 1681 GlyLeuGlnAlaAspSerLeuArgLeuGluProGlyArgAlaAsnAapGlyAapTrpHis 1700  
 DB 5041 GGGCTTCAGGGCCCTCTCTCTGAGGAGCCAGCGGGGCCAATACCGGTGACTGGCCAC 5100  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAap 1720  
 DB 5101 CATGCAAGCTGGCATGGAGGCCAGGGGGGAGCTGGCCATTCCTGTCTTCCAT 5160  
 QY 1721 TyrrGlyGlnGlnArgAlaGluGlyAenLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 DB 5161 TATGGCAGAGAGAGACAGAGGGCACTGGGCCCCCGGCTGTGATGTCTGACACTGAGC 5220  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760

DB 5221 AACATTAAGTGGCGGAAATACCTGGGCAAGCCGGGTGGGCCCCGGCTTTCGGGGCC 5280  
 QY 1761 CysLeuGlnGlyValArgValSerAapThrProGluGlyValAsnSerLeuAapProSer 1780  
 DB 5281 TGTGTGACGGGTGTGGGTGAGGATACCGCAAGGGGGTTAAACAGCTGGAATCCAGC 5340  
 QY 1781 HisGlyLysSerIleAsnValGluGlnGlyCysSerLeuProAapProCysAapSerAsn 1800  
 DB 5341 CATGGGAGAGACATCAACGTGGACCAAGGCTGTAGCTGAGCCCTGTGACTCAAC 5400  
 QY 1801 ProCysProAlaAsnSerTyrrCysSerAsnAapTrpAapSerTyrrSerCysSerCysAap 1820  
 DB 5401 CCGTGTCTGTCAAGCTATTGCAAGCAACAGCTGGGCAAGCTATTCTCGACGCTGAT 5460  
 QY 1821 ProGlyTyrrGlyLysAapAsnCysThrAapValCysAapLeuAapProCysGluHisGln 1840  
 DB 5461 CAGGTTACTATGGTGAACCTGTAATGTGTGACCTGAAACCGGTGAGACACAG 5520  
 QY 1841 SerValCysThrArgLysProSerAlaProHisGlyTyrrThrCysGluCysProProAsn 1860  
 DB 5521 TCTGTGTATACCCGACAGCCAGTGGCCCCCATGGCTATACCTGGAGTGTCCCCAAT 5580  
 QY 1861 TyrrLeuGlyProTyrrCysGluThrArgIleAapGlnProCysProArgGlyTrpTrpGly 1880  
 DB 5581 TACCTGGGCGATCTGTGAACACAGATTTGACAGGCTTGTCCCGGTGGTGGGGGA 5640  
 QY 1881 HisProThrCysGlyProCysAapCysAapValSerLysGlyPheAapProAapCysAap 1900  
 DB 5641 CATCCCATGTGGCCCATGGCACTGTGATGTACAGAAAGGCTTTGACCCAGACTGACAC 5700  
 QY 1901 LysThrSerGlyGlyCysHisCysLysGlyLysAsnHisGlyArgProProGlyLysProThr 1920  
 DB 5701 AAGACAGCGGCGAGTGGCACTGCAAGGAGAACCACTACCGGCCCGCAGGACGCCAAC 5760  
 QY 1921 CysLeuLeuCysAapCysTyrrProThrGlySerLeuSerArgValCysAapProGluAap 1940  
 DB 5761 TGCCCTGTGTGACTGTATCCCAAGGCTCTCTGTCACAGTCTGTGACCCGAGAGAT 5820  
 QY 1941 GlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAapArgCysAapAapPro 1960  
 DB 5821 GGCAGTGTCAATGCAACCAAGGTGTCACTGGGGGTCAAGTGTACCCGTGTGACACCT 5880  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGlyValAsnTyrrAapSerCysProArgAlaIle 1980  
 DB 5881 TTGCTGAGTTCACCAACCAATGGCTGTGAAGTGAATTATGACAGTGGCCACAGCGAAT 5940  
 QY 1981 GlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
 DB 5941 GAGGTGGGATCTGGTGGCCCCGTACCGCTTGGGCTGCTGCTGCTGCTGCTGCTCC 6000  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAapGlyHisArgGlyTyrrLeuProPro 2020  
 DB 6001 AAGGCTTTTGGGAGCTGTGTGCGCCACTGTATGAGCAGAGGGGTGGCTCCCCCA 6060  
 QY 2021 AenLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu 2040  
 DB 6061 AACCTTCACTGACAGTGCATCACTTCTCAAACTGAAGGGCTTGGCTGAGAGGGCTA 6120  
 QY 2041 GlnArgAsnGlnSerGlyLeuAapSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg 2060  
 DB 6121 CAGCGGAATGAGTCAAGGCTTGAAGCTCAGGGGGCTCCCAAGAGCTAGGCTCTCTGGGC 6180  
 QY 2061 AsnAlaThrGlnHisPheThrAlaGlyTyrrPheGlySerAapValLysValAlaTyrrGlnLeu 2080  
 DB 6181 AACGCAAGCAGCACAAGCTGTGCTACTTGGGAGCAGACGTCAAGTGGCTTACCAAGCTG 6240  
 QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 DB 6241 GCCAGCGCGTGTGGCCCAAGAGACCCAGAGGGGGCTTGGGCTGTGTGCAACAG 6300  
 QY 2101 AapValHisPheThrGlnAenLeuLeuArgValGlySerAlaLeuLeuAapThrAlaAsn 2120



D	6301	GACGTGCACCTTCACTGAGAAATCTGCTGGCGGTGGGCAAGCGCCCTTCTGGACAAGCCAAAC	6360
O	2121	LYSAAGNISTPGLU <sup>LEU</sup> LEGLINGL <sup>THR</sup> GLUGLYGL <sup>THR</sup> ALATP <sup>LEU</sup> GLNHIS	2140
D	6361	AAGGGCACTGGAGACTGATCCAGACAGACAGAGGGTGGCACCGCTGGCTCTCCAGAC	6420
O	2141	TYRGLUALATYALASER <sup>ALA</sup> LEUALAGLNANNE <sup>LARGH</sup> STHYR <sup>LEU</sup> SER <sup>PRO</sup> PE	2160
D	6421	TATGAGGCTTACGCCAGTGGCCCTGGCCCAAGAACTGCGGCACACCTTAACCTAAGCCCCCTTC	6480
O	2161	THRILEVAL <sup>THR</sup> PRO <sup>ASN</sup> ILEVALILESER <sup>VAL</sup> VALAGLEA <sup>ASP</sup> LYSGLY <sup>ASN</sup> PH <sup>ALA</sup>	2180
D	6481	ACCACTGTCACGGCCCAACAATGTGCATCTCC <sup>TA</sup> AGGCCCTTGGACAAAGGAAACCTTGGCT	6540
O	2181	GLYALATY <sup>LEU</sup> PRO <sup>ARG</sup> TYRGLUALALEU <sup>ARG</sup> GLUGLYGLINGL <sup>PRO</sup> PRO <sup>ASP</sup> LEU <sup>GL</sup> THR	2200
D	6541	GGGGCCAAAGCTGGCCCCGCTACGAGGCCCTGGGTGGGAGACAGCCCCCGA <sup>CTT</sup> GAAGACA	6600
O	2201	THRVALILEU <sup>PRO</sup> GL <sup>USER</sup> VALPHEARGGL <sup>THR</sup> PRO <sup>VAL</sup> VALARG <sup>PRO</sup> ALAGLY	2220
D	6601	ACAGTCATTTGCGCTGAGCTGTCTTTCAGAGAGACGCCCCCGTGTACGGCCCGGACGC	6660
O	2221	PROGLYGLUALAGL <sup>ING</sup> L <sup>PRO</sup> GLUGLYLEUALAARG <sup>ARG</sup> L <sup>ARGH</sup> IS <sup>PRO</sup> GL <sup>LEU</sup>	2240
D	6661	CCCCGAGAGGGCCCAAGAGACAGAGAGCTGGCAACGGCAACGCAACGGCACCCGAGACTGG	6720
O	2241	SERGL <sup>ING</sup> LYGLUALAVALALASER <sup>VAL</sup> ILEILEY <sup>TRARG</sup> THRLEUALAGLY <sup>LEU</sup> LEU <sup>PRO</sup>	2260
D	6721	AGCCAGGGTGAAGGCTGTGGCCAGGCTCATCATCTACCGCACCCCTGGCGGGCTACTGCT	6780
O	2261	HISANTY <sup>TRASP</sup> PRO <sup>ASP</sup> LY <sup>ARG</sup> SER <sup>LEU</sup> ARGVAL <sup>PRO</sup> LY <sup>ARG</sup> PROILE <sup>ASN</sup> THR	2280
D	6781	CATAACTATGACCCCTGACAGCGCAGCTGGAGAGTCCCAACCCCGCATCATGACACA	6840
O	2281	PROVAL <sup>VAL</sup> SER <sup>ILE</sup> SER <sup>VAL</sup> HIS <sup>ASP</sup> APGLUGLYLEU <sup>LEU</sup> PRO <sup>ARG</sup> ALALEU <sup>ASP</sup> LYS	2300
D	6841	CCCCGTGGAGCAATCAGGCTCCATGATGATGAGAGGCTTCTGCCCGGGCCCTGACAACAA	6900
O	2301	PROVAL <sup>THR</sup> VALGL <sup>INP</sup> HE <sup>ARG</sup> LEU <sup>LEU</sup> GL <sup>ING</sup> GLUGLYUALAG <sup>THR</sup> LYS <sup>PRO</sup> ILECY <sup>VAL</sup>	2320
D	6901	CCCGTCACGGTGCAGTTCCGCTGTGTGAGACAGAGAGCGGACCAAGACCCATCTGTGTC	6960
O	2321	PHE <sup>TRP</sup> ASN <sup>ISER</sup> ILEU <sup>VAL</sup> SERGL <sup>Y</sup> THRGLY <sup>TP</sup> SER <sup>ALA</sup> ARGGL <sup>Y</sup> CYGLU	2340
D	6961	TTCTGGAAACCATTCATCTGCTGTCAGTGGACAGATGGCTGTGGCCAGAGGCTGGAA	7020
O	2341	VALVALPHE <sup>ARG</sup> ANGL <sup>USER</sup> HIS <sup>VAL</sup> SERCYGL <sup>NCY</sup> ASN <sup>ISW</sup> ETH <sup>SR</sup> PE <sup>HALA</sup>	2360
D	7021	GTCGTCTTCCGC <sup>AA</sup> TGAGGCAAGCTCAGCTCCAGTGCACAACACATGACGAGCTTGCT	7080
O	2361	VALLEU <sup>MSER</sup> VALSER <sup>ARG</sup> ARGGLUGLYL <sup>ILEU</sup> PRO <sup>LEU</sup> LYSTHR <sup>LEU</sup> THR	2380
D	7081	GTCGTCAN <sup>GAC</sup> GTTTCTCGCGGGAGAAATGGGAGATCTCGGCACATGAAAGACATGACA	7140
O	2381	TYRVALALALEUGLY <sup>VAL</sup> THR <sup>LEU</sup> ALALEU <sup>SEU</sup> THR <sup>PH</sup> EPHE <sup>LEU</sup> THR <sup>LEU</sup>	2400
D	7141	TACGTGGCTCAGATGTCACCTTGGCTGGCCCTTCTGCTCACCTTCTTCTTCCACCTTC	7200
O	2401	LEUARG <sup>ILE</sup> LEUARG <sup>SE</sup> ASN <sup>GLNHIS</sup> GLYILEARG <sup>ARG</sup> ANLEU <sup>TRAL</sup> ALALEU <sup>GLY</sup>	2420
D	7201	TTGGGTATCCGCTCGCTCCAAACCAACGAGCATCCGATACATGACAGCTGGCCCTGGGC	7260
O	2421	LEUALAG <sup>LEU</sup> VALPHE <sup>LEU</sup> LEU <sup>GLY</sup> ILEASN <sup>GL</sup> ALASP <sup>LEU</sup> PRO <sup>PH</sup> ALAC <sup>Y</sup> THR	2440
D	7261	CTGGCTACAGCTGGCTTCTCTCTGGAAATCAACAGAGCTACCTCCCTTTGGCTGACACA	7320
O	2441	VALILEALALEU <sup>LEU</sup> HIS <sup>PH</sup> LEU <sup>TYR</sup> LEUCY <sup>THR</sup> PHESER <sup>TRAL</sup> ALALEU <sup>GLU</sup>	2460
D	7321	GTCATTCGCANCCSGCTGCACCTTCTGTAACCTCTGCACCTTTTCTGGGCTCTGCTGGAG	7380
O	2461	ALALEU <sup>HIS</sup> LEU <sup>TYR</sup> ARGALALEU <sup>THR</sup> GLUVALARG <sup>ASP</sup> VAL <sup>ASN</sup> THRGLY <sup>PRO</sup> ME <sup>LARG</sup>	2480
D	7381	GCTTTGACCTGTACCGGGCACTCACTGAGGTGGCATATGACAACCGGCCCCCATGCGC	7440

QY	2481	PheTyrTrpMetLeuGlyTyrTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Db	7441	TTCTACTACATGCTGGGCTGGGGGTGGCTGGCTTCAATCACAGGAGCTAGACGGTGGCTGG	7500
QY	2501	AspProGluGlyTyrGlyValAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle	2520
Db	7501	GACCCCGAGGGCTACGGGAAACCTGACTTCTGTGGCTTCCACTATGACACGGCTATC	7560
QY	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu	2540
Db	7561	TGGAGTTTGGCTGGGCCGGTGGCTTGGCCGTCCGATGAGTGTCTTCCCTGTACATCCTG	7620
QY	2561	AlaAlaAspAlaSerCysValAlaGlnArgGlnIlyPheGlyIlyGlyGlyProValSer	2560
Db	7621	GCGGCGCCGGGCTCTCTGTGGTCCACAGGGAGGAGCTTGGAAAGAAAGTCCGTCTCG	7680
QY	2561	GlyLeuGlnProSerPheAlaValIleLeuLeuLeuSerAlaThrTrpLeuAlaIleu	2580
Db	7681	GGCCTGAGACCTCTCTTGCGGCTCTCTGTGTGTGAGCGGCACGTGGCTGTGGACTG	7740
QY	2581	LeuSerValaSerAspThrIleuLeuPheHisTyrIleuPheAlaThrCysAsnCysIle	2600
Db	7741	CTCTGTGCACACAGCAGACCCCTCTCTTCCACTACCTTTTGCTACTGTCAATTGGATC	7800
QY	2601	GlnGlyProPheIlePheLeuSerTyrValIleuSerIlyGlyValAlaGlyValaIleu	2620
Db	7801	CAGGGCCCCCTTCACTTCTCTCTCTTAATGTGGTCTTACAGAGAGGTCGGAAAGCACTC	7860
QY	2621	LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrIlySerThrLeu	2640
Db	7861	AAGCTTGGCTGCACCGGACAGCCAGCCGTGACCTGTCTGACCAACCAAGTCCACCTG	7920
QY	2641	ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly	2660
Db	7921	ACCTGTCTTACACTGTCCCAAGCCCTTACGCAATGGGGGCTGACAGCCCTTAAGA	7980
QY	2661	AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle	2680
Db	7981	GACTCGCGGCTCTCTGCACAGCACAGTGTGCTCGGGCAGAAAGTCAAGCCAGTACATC	8040
QY	2681	ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly	2700
Db	8041	CCCTTCTTGTGTGAGGGAGGAGTCCGCACTGAACCTTGCCAAAGGCCCCCTGGCTGGGG	8100
QY	2701	AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer	2720
Db	8101	GATCCAGGACCTGTGTCTGTGAAGGTTCAGACCAAGCATGATCTTGCACAGGACTCC	8160
QY	2721	AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer	2740
Db	8161	GACAGTACCTGTCTTAAAGAAAGACCAAGTGGGCTCTATGGCTTACCACTCATCA	8220
QY	2741	AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGly	2760
Db	8221	GACAGTAAAGAGAAAGAGAGAGAGAGAGAGAGAGGCCCTTCCCTGTGAAGACAGGGC	8280
QY	2761	TrpAspSerLeuLeuGlyProGlyAlaGlnArgLeuProLeuHisSerThrProLysAsp	2780
Db	8281	TGGGATAGCCTGTCTGGGGCTGTGAAGCAGAGATGCTCCCTGCACAGTACTCCCAAGAT	8340
QY	2781	GlyGlyProGlyProGlyIlyValaProTyrProGlyAspPheGlyThrThrAlaIlyGln	2800
Db	8341	GGGGGCCCAAGGGCTGTGCAGAGGCCCTCTGGCCAGAGACTTGGGACACACAGCAAAAG	8400
QY	2801	SerSerGlyAsnGlyAlaProGlnGlnArgLeuArgLeuGlnAsnGlyAspAlaLeuSerArg	2820
Db	8401	AGTATGGCAACGGGGCCCTTGAAGACGGCTGTGGGAGATGAGATGAGTCCCTGTCTCGA	8460
QY	2821	GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisIlyGly	2837
Db	8461	GAGGGGTCCCTAAGGCCCTTCCAGGCTCTTGTGCCAGGCTCACAAAGGT	8511





Db	1839	AGCCCTGGTAACTTAAAGACTACGAGGACCGGCGCTGAGTAAGCTCTGGCATCCAGGCGAC	1898
QY	999	rSerLlaProLeuValSerArgAlaThrValHisValArgLeuLeuMpaArgAnaPaa	1019
Db	1899	GTACGCTCTCTGGGAGCGGGGCTACAGTCACAGTCGCGCTCTTGAACGGCAATGACAA	1958
QY	1019	nProProValLeuGluYasnPheGluIleLeuPheAsnAsnTyValThrAsnArgSer	1039
Db	1959	CCACACAGATGCTGGGGCAACTTTAGAGATCCTTTCAACACATATCAACATCCCTCAG	2018
QY	1039	rSerPheProGluYglYalIleGlyArgValProAlHisAspProAlIleSerAspSe	1059
Db	2019	CAGCTTCCCTGGGGGTGGCATTTGGCCGAGTACCTGGCCATACCTGATATCTCAGATAG	2078
QY	1059	rLeuThrTyrrSerPheGluNalArgGlyYasnGluLeuSerLeuValLeuLeuAsnAlaSerTh	1079
Db	2079	TCTGACTTAAAGCTTTAGACGGGGAAATGAATCACTAGCGTGCTCTGCTCAATGCTTCAC	2138
QY	1079	rGlyLeuLeuYlsLeuSerArgAlaLeuAspAsnAsnArgProLeuGluNalIleMetSe	1099
Db	2139	GGGTAGCTGAAGCTTAAGCCCGGCACTGGACAAACAACCGGCTTGGAGGCCATCAAG	2198
QY	1099	rValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrI	1119
Db	2199	CGTGTGGGTGCAGACGGCGGTACACAGCGTGAACGGCCAGTGGCGGCTGTGGTGAACAT	2258
QY	1119	AlleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProG	1139
Db	2259	CATACCGATAGATGTCAACCCACACATCAACCTTCGCGCTGGAGGACATGTCAACCGCA	2318
QY	1139	nArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaTh	1159
Db	2319	GGCGTCTCTGCACACACTGCTAGGCTCTTCAATCAAGCGGTGGCCGCAACGCTGGCCAC	2378
QY	1159	rProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGluYglYH	1179
Db	2379	GCCACCGGACCACTGGTGGTCTTCAACGTACAGCGGACACCGAGCGCCCGGGGGCCA	2438
QY	1179	AlleLeuAsnValSerLeuSerValGlyGlnProProGlyProGluYglYProProPh	1199
Db	2439	CATCTTCAACCTGATGAGCTGTGCGGTGGGCGACGGCCGCGGGGCGAGCGCGCTT	2498
QY	1199	eLeuProSerGluAsnLeuGlnGluNalGlyLeuTyrrLeuAsnArgSerLeuThrAlaI	1219
Db	2499	CTGCGCTCTAGGAACTTGGACGAGGCGCTTAATCTCAACCGAGCTGCTGACGGCCAT	2558
QY	1219	eSerAlaGlnArgValLeuProPheAspAsnAlleCysLeuArgGluProCysGluAs	1239
Db	2559	CTCGGCAACAGCGGTGTGGCTTTCGACGACCAACATGTGCTGGGGAGGCGCTGGAGAA	2618
QY	1239	nTyrrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSe	1259
Db	2619	CTACATGCGCTGCGCTGTGCGGTGCTGCGCTTCGACTCTCCGCGGCTTCAATCGCTCTC	2678
QY	1259	rSerValLeuPheAspProIleHisProValGlyYleuArgCysArgCysProProG	1279
Db	2679	CTTCGTGCTCTTCGGGCCAATCCACCCCGTGGAGGGCTGCGCTGCGCGCGCGCGG	2738
QY	1279	yPheThrGlyAspTyrrCysGluThrGluValAspLeuCysTyrrSerArgProCysGlyPr	1299
Db	2739	CTTCAACGGGTGACATACGACGAGAACCGAGGTGAGACTCTGTGTAATCGCGGCGCTGTGGCCC	2798
QY	1299	OhIleGlyArgCysArgSerArgGluGlyTyrrThrCysLeuCysArgAspGlyTyrrTh	1319
Db	2799	CCAGGGGCGCTGCGCGACCGCGAGGGCGGCTACACTGCTCTGTCTGTAATGACTACAC	2858
QY	1319	rGlyGlyHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIysAs	1339
Db	2859	GGGTAGACACTGTAGGTGAGTGTGCTGACTAGGCGGTGGACCCCGGGGTGTCTGGAGAA	2918
QY	1339	nGlyGlyThrCysValAsnLeuLeuValGlyYleuTyrrCysAspCysAspProSerGlyAs	1359

Db 2919 TGGGGGACCTGTGTCACCTGCTGGTGGCGGTTTCAAGTGCATTCGACATTCGAGAA 2978  
 QY 1359 pPheGlyuysProGlyrCyseGlnValThrThrArgserRheProAlaHisSerPhe1Leh 1379  
 Db 2979 CTTCGAGAAAGCCCTTACGACGGTGAACGCGGACCTTCCCGCCACTCTTCATCAAC 3038  
 QY 1379 rPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLeuG1 1399  
 Db 3039 CTTTGGCGGCGCTGGCGCAAGCTTTCACCTTCACCTCGGCCCTCTCGTTGGCCAAAGAA 3098  
 QY 1399 uArgAspGlyLeuLeuLeuLeuTyraSnglyArgPheAsnglyArgPheAsnglyVala 1419  
 Db 3099 GGGCGACGGGTGTGCTGTGTACATGGGGCTTCAATGAGAACATGATGATTTGGCCCT 3158  
 QY 1419 uGlnVal111eGlnGlnGlnVal1GlnLeuThrPheSerAlaGlyGlnSerThrThrVa 1439  
 Db 3159 CAGAGTGAATCCAGAACAGAGTCCAGCTCACTCTCTGACGGGGAGTCAACCAACGAGT 3218  
 QY 1439 lSerProPheValProGlyGlyValSerAspGlyGlnTrpHisThrValGlnLeuLeuTy 1459  
 Db 3219 GTCCCAATTCGTGGCCGGAAGTCAGTATGGCCAGTGGCAATACGCTGCACTGAATA 3278  
 QY 1459 rTyraSnglyProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnLeuVa 1479  
 Db 3279 CTACAAATAGCCACTGTGTGGTCAAGACAGGGCTCCCAAGGGCCCATCAGACAGAAAGGT 3338  
 QY 1479 lAlaValValThrValAspGlyCyseAspThrGlyValAlaLeuArgPheGlySerValle 1499  
 Db 3339 GGGCTGTGTGAACCGTGAATGGCTGTGAACAGAGTGGCCGTGGATCTGTCT 3398  
 QY 1499 uGlyAsnTyraSerCyseAlaAla-GlnGlyThrGlnGlyGlySerGlySerLeuAspL 1519  
 Db 3399 GGGCAACTACTCTGTGTGTGCCCAAGGACCAAGGGTGGCAGCAAGAACTCTCGAATC 3458  
 QY 1519 eutThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlyLeuSerPheProValaGm 1539  
 Db 3459 TAAAGGGGGCCCTGTCTACTAGAGGGGGTGTCTGACCTGCCCAAGAACTTCCAGTCCGA 3518  
 QY 1539 eArgGlnPheValGlyCyseMetArgAsnLeuGlnValAspSerArgHisLeuMetA 1559  
 Db 3519 TGGGGAGATTCTGTGGGCTGCATTCGGAACTCGAGGTGAACGCGGCACTAGACATGG 3578  
 QY 1559 lAspPhe1LeuAlaAsnAsnglyThrValProGlyCyseProAlaGlyValAsnValCyse 1579  
 Db 3579 CTGACTTCACTTGCACAAATGGCAACGTCCTGGCTGCCCTGCCAAGAAAGAACTGTGTG 3638  
 QY 1579 gSerSerAsn-ThrCyseHisAsnGlyGlyThrCyseValAsnGlnThrAspAlaPheSerCyse 1598  
 Db 3639 ACAGCAA-AACCTTGCACAAATGGGGCACTTGTGTAAACCAAGTGGACCGCTTCAGCTGC 3697  
 QY 1599 GlnCyseProLeuGlyPheGlyGlyTyraSerCyseAlaGlnGlnMetAlaAsnProGlnHis 1618  
 Db 3698 GAGTGGCCCTTGGGCTTGGGGGCAAGAGCTGGCCCAAGAAATGCCCAATCCAGCAC 3757  
 QY 1618 PheLeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProL1eserGlnProTrp 1638  
 Db 3758 TTCCTGGGACGACGCTGTGTGGCTGCATGGCTCTCGCTGCCATCTCCCAACCTG 3817  
 QY 1639 TyraLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAla1LeuThr 1658  
 Db 3818 TACCTAGGCTCAATGTTCCGACGCGCCAGCGAGGTCTGTGTGACGGCAATCAC 3877  
 QY 1659 ArgGlyArgSerThr1LeuThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGln 1678  
 Db 3878 AGGGGGCGAGCAACCACTTACAGCTTACAGAGGGGCAAGTATGCTGAGCGTGGAG 3937  
 QY 1679 GlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAsp 1698  
 Db 3938 GGGACAGGGCTTACGGCTCTCTCTCGCTGAGACAGCGCGGCGCAATGACGTGTAC 3997  
 QY 1699 TrpHisAlaGlnLeuAlaLeuGlyAlaSer-GlyGlyProGlyHisAla1LeuSer 1718  
 Db 3998 TGGCACTACGACAGCTGGCACTGGAGGCAAT-CGGGGGGCTTGGCCATGCCATCTGTG 4056

QY 1718 rPheAspTyraGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHis 1738  
 Db 4057 CTTCATTAATGGGACGAGAGAGAGAGGAGCAACTGGGCCCCCGGCTGCATGGCTGTGA 4116  
 QY 1738 sLeuSerAsn1LeuThrValGlyGly1LeProGlyProAlaGlyGlyValAlaArgGlyPh 1758  
 Db 4117 CCTGAGCAACATTAACATGGGGCGGAAATACCTGGGCGAGCGGGGTGTGGCCGTGCTT 4176  
 QY 1758 eArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValaAsnSerLeuAs 1778  
 Db 4177 TGGGGCTGTGTGACGGGTGTGGCGGTGAAGCAATACGCAAGGGGGTTAAACGCTTGA 4236  
 QY 1778 pProSerHisGlyGlnSer1LeuAsnValGlnGlnGlyCyseSerLeuProAspProCyseAs 1798  
 Db 4237 TCCAGCCATGGGAGAACATCAACGTGAGAACAGGCTGTAGCTGCTGACCTGTGTGA 4296  
 QY 1798 pSerAsnProCyseProAlaAsnSerTyraCyseSerAsnAspTrpAspSerTyraSerCyse 1818  
 Db 4297 CTCAACCCGTCCTGTCTTACACTATTCAGCAACGACTGGGACAGCTATTCCTGACG 4356  
 QY 1818 rCyseAspProGlyTyraTyraGlyAspAsnCyseThrAsnValCyseAspLeuAsnProCyseG1 1838  
 Db 4357 CTGTGATCCAGGTTACTATGTGTGAACACTGTACTAATGTGTGACCTGAACCCGTGTGA 4416  
 QY 1838 uHisGlnSerValCyseThrArgGlyProSerAlaProHisGlyTyraThrCyseGlnCysePr 1858  
 Db 4417 GCAACAGCTGTGTGTACCCGCAAGGCCAGTGGCCCCCATAGGCTTATCTGAGATGCC 4476  
 QY 1858 oProAsnTyraLeuGlyProTyraGlyuThrArg1LeAspGlnProCyseProArgGlyTy 1878  
 Db 4477 CCAAAATTAACCTTGGGCAATCTGTGAGAACAGGATTTGAACAGGCTTGTCCCCGGTGTG 4536  
 QY 1878 pTrpGlyHisProThrCyseGlyProCyseAsnCyseAspValSerGlyGlyPheAspProAs 1898  
 Db 4537 GTGGGACATCCACATGTGGCCCAATGCATGTGAATGACGAAGGCTTTCACCA 4596  
 QY 1898 pCyseAsnTyraSerGlyGlnGlyCyseHisCyseGlyuAsnHisTyraArgProProGlySe 1918  
 Db 4597 CTGCAACAAACAGAGCGGCGAGTGCATCTGCAAGAGAAACCACTACCGGCCCAAGGAG 4656  
 QY 1918 rProThrCysLeuLeuCyseAspCytyrProThrGlySerLeuSerArgValCyseAspPr 1938  
 Db 4657 CCCCACCTGCTCTGTGTGTGACTGTACCCCAAGGCTCTTGTCTCAGAGCTGTGAACC 4716  
 QY 1938 oGlnAspGlyGlnCyseProCyseGlyProGlyVal111eGlyArgGlnCyseAspArgCyseAs 1958  
 Db 4717 TGAAGATGGCAAGTGTCCATGACAGCGAGGTGTCAAGGGGCTCAAGTGTGAACCGCTGTGA 4776  
 QY 1958 pAsnProPheAlaGlnValThrThrAsnGlyCyseGlnValaAsnTyraAspSerCyseProAr 1978  
 Db 4777 CAACCTTTTCTGAGGTCAACCAACATGGCTGTGAAGTGAATTAATGACACTGCCACG 4836  
 QY 1978 gAla111eGlnAlaGly111eTrpProArgThrArgPheGlyLeuProAlaAla1a1aPr 1998  
 Db 4837 AGCAATTAAGCTGGGAATCTGTGGCCCGTAAACCGCTTCCGGGCTGTGTGCTGCC 4896  
 QY 1998 oCyseProGlySerPheGlyThrAlaValaArgHisCyseAspGlnHisArgGlyTyraPhe 2018  
 Db 4897 CTGTCCAAAGGCTCTTGGGACTGTGTGCGGCACTGTATGAGCACAGGGGGTGTCT 4956  
 QY 2018 uProProAsnLeuPheAsnCyseThrSer1LeuThrPheSerGlnLeuLeuGlyPheAlaG1 2038  
 Db 4957 CCCCCCAACCTTCAACTGACAGTGCATCACTTTCAGAACTGAAGGGCTTGTGCTGA 5016  
 QY 2038 uArgLeuGlnArgAsnglyuSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLe 2058  
 Db 5017 GCGGTACAGGGGAATGTAGTGAAGCTTACACTCAGGGGGCTTCCAGCAGCTAGCGCTGT 5076  
 QY 2058 uLeuArgAsnAlaThrGlnHisThrAlaGlyTyraPheGlySerAspValTyraVala1aTy 2078  
 Db 5077 CTTGGCAACGCAACGACGACACAGCTGGCTTACTTGGGACGACAGCTCAAGGTGGCTTA 5136

QY 2078 rglndleuAlaThrArgleuLeuAlaHISgluSerThrThrArglyPheglYleuSerAl 2098  
 Db 5137 CCAAGTGGCCACAGCGCGCTGGCCACAGAGCACCCAGCGCGCTTGGGCTGTCTGCG 5196  
 QY 2098 aThrglnAspValHISpHeThrGluAsnLeuLeuArgValglYserAlaLeuLeuAspTh 2118  
 Db 5197 CACACAGAGCGTGCATTCACCTGAGAAATCTGCTGGGGTGGGACAGCGCTCTCGAGAC 5256  
 QY 2118 rAlaAsnLysArgHISrTPGluLeuIleGlnGlnThrGluGlyYThrAlaATrPLeuLe 2138  
 Db 5257 AGCCAAACAGCGCGCTGGAGCTGATCCAGCAGACAGAGGGGTGCACCGCTGCTCTCT 5316  
 QY 2138 uGlnHISrYrGlnuAlaYrAlaASerAlaLeuAlaGlnAsnMetArgHISrThrYrLeuSe 2158  
 Db 5317 CCAGACTATATAGGCTATAGCGCAGTGGCCCTGGCCCAACATGCGGCACACTTACTTAA 5376  
 QY 2158 rProPheThrIleValThrProAsnIleValIleSerValValArgLeuAspLysGlyAs 2178  
 Db 5377 CCCCTTCAACATCGTCACGCCCACACATTTGTCACTCCGTAATGCGCTTGGACAAAGGGA 5436  
 QY 2178 nPheAlaGlyAlaLysLeuProArgYrGluAlaLeuArglyGlyGlnProProAspLe 2198  
 Db 5437 CTTTGCTGGGGCCAAAGCTGCGCCCGCTACAGGCGCTGCTGGGGAGAGCGCCCGGACCT 5496  
 QY 2198 uGlnThrThrValIleLeuProGlnuSerValPheArgGluThrProProValValArgPr 2218  
 Db 5497 TGAACAAACATGATCTGCTGCTGAGTCTGCTTCAAGAGAGAGCGCCCGCTGCTCAGGCGC 5556  
 QY 2218 oAlaGlyProGlyGluAlaGlnGluProGlnuLeuAlaArgArgGlnArgArgHISPr 2238  
 Db 5557 CGAGAGCCCCCGAGAGCGCCACAGAGCCAGAGAGCTGGCAGCGGACAGCCACGACCC 5616  
 QY 2238 cGluLeuSerGlnGlyGluAlaValAlaSerValIleIleYrArgThrLeuAlaGlyLe 2258  
 Db 5617 GGAACCTGAGCGAGGGTGAAGCTGTGGCCAGCGTCATCATCCGACACCTGGCGCGGCT 5676  
 QY 2258 uLeuProHISerYrAspProAspLysArgSerLeuArgValProLysArgProIleI 2278  
 Db 5677 ACTCCCTCATATACATGACCTTGAACAGCGACTTGAAGTCTCCCAAAACCCGATCAT 5736  
 QY 2278 eAsnThrProValIleSerIleSerValHISAspAspGluGluLeuLeuProArgAlaLe 2298  
 Db 5737 CAACACACCCCTGGTGAAGCATCACGCTCCATGATGATGAGAGCTTCTGCCCCGGGCTCT 5796  
 QY 2298 uAspLysProValIleThrValGlnPheArgLeuLeuGlnuThrGluuArgThrLysProI 2318  
 Db 5797 GGAACAAACCCCTGACGGGTGAGTTCGCGCTGCTGAGACACAGAGGCGGACCAAGCCCAT 5856  
 QY 2318 eCyValPheThrAsnHISerIleLeuValSerGlyYrThrGlyYrTPSerAlaArgG 2338  
 Db 5857 CTGTGTCTTCTGGAACCAATTCATCTGTGTCAATGAGCAACAGTGGCTGTGCGCAGAGG 5916  
 QY 2338 YCyGluValValPheArgAsnGlnuSerHISValSerCyAsnGlnCyAsnHISMetThrSe 2358  
 Db 5917 CTGTGAAGTCTCTTCCGCAATGAGAGCCATGACGCTGCGACAGTGCACATGAGAGAG 5976  
 QY 2358 rPheAlaValLeuLeuAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysTh 2378  
 Db 5977 CTTGCTGTGCTCATGAGCGTTTCTCGCGGAGAAATGGGAGATCTCGACATGAAAGAC 6036  
 QY 2378 rLeuThrThrValAlaLeuGlyValIleThrLeuAlaAlaLeuLeuLeuThrPhePhePheLe 2398  
 Db 6037 ACTACACATACCTGGCTCTTAAAGTGTACCTTGGCGCCCTTCTGTCTCACCTTCTTCTCT 6096  
 QY 2398 uThrLeuLeuArgIleLeuArgSerAsnGlnHISglYIleArgArgAsnLeuThrAlaAl 2418  
 Db 6097 CACTCTCTGGGATTCCTGGCTCCAAACCAACAGGCAATCCGATTAACCTGACAGCTGCG 6156  
 QY 2418 aLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAl 2438  
 Db 6157 CTTGGGGCTGGCTCAGGCTGTCTTCTCTGAGATCAACAGGCTGACCTCCCTTTTTCG 6216  
 QY 2438 aCyThrValIleAlaIleLeuLeuHISpHeLysrYrLeuCyThrPheSerTrpAlaLe 2458

Db 6217 CTGCACAGTCAATGCCATCTGCTGACATTCCTGTACTCTGACACTTCTGGGCTCT 6276  
 QY 2458 uLeuGlnAlaLeuHISerYrArgAlaLeuThrGluValArgAspValAsnThrArgLysPr 2478  
 Db 6277 GCTGAAGGCCCTGACCTGTACCCGGGACATCACTGAAGGTGGCAATGTCAACACCGGCC 6336  
 QY 2478 oMetArgPheYrYrMetLeuGlyTrPGlyValProAlaPheIleThrGlyLeuAlaVala 2498  
 Db 6337 CATCGCTTCTACTACATAGCTGGGCTGGGGGTGCTGCTTATTCACAGGGCTAGCCGT 6396  
 QY 2498 lGlyLeuAspProGlnuGlyYrGlyYAsnProAspPheCySerTrpLeuSerIleYrAspTh 2518  
 Db 6397 GGGCTGAGCCCGGAGGGCTTACGGGAACCTGACTTGTGCTGCTCTTCACTATGACAC 6456  
 QY 2518 rLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuY 2538  
 Db 6457 GCTCATCTGAAGTGTCTGACCGGTGGGCTTGTGGCCGTGCTCATATGAGTGTCTTCTGT 6516  
 QY 2538 rIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGlnuLysLysGlyPr 2558  
 Db 6517 CATCTGGCGGCCCGGCGCTCTCTGTGTGCTGCCACGCGGCAAGGCTTGAAGAAAGTCTC 6576  
 QY 2558 oValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLe 2578  
 Db 6577 TGTCTCGGGCTGACGCTCTTCCGCGCTGCTCTGCTGCTGAGCCGACAGTGGCTGCT 6636  
 QY 2578 uAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHISrYrLeuPheAlaThrCyAs 2598  
 Db 6637 GGCATCTGCTCTGTCAACAGCAACCTTCTCTTCACTACCTCTTCTCACTGCA 6696  
 QY 2598 nCyAlleGlnGlyProPheIlePheLeuSerYrValValLeuSerYsGluValArgLy 2618  
 Db 6697 TTGATCAAGGGCCCTTATCTTCTCTCTCTATGTGTCTTACAGAGAGGTCCGGA 6756  
 QY 2618 sAlaLeuLysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSe 2638  
 Db 6757 AGACATCAACCTTGCCTGCGCGGCAAGCCACAGCCCTGACCTGTCTGACCAACAGTC 6816  
 QY 2638 rThrLeuThrSerSerYrAsnCySProSerProYrAlaAspGlyYArgLeuYrGlnPr 2658  
 Db 6817 CACCTGACCTCTGCTCTTACATCTGCCCCAGCCCTTACAGAGATGGGCGCTGATCACGCC 6876  
 QY 2658 oYrYr-GlyAspSerAlaGlySerLeuHISerThrSerArgSerGlyLysSerGlnPro 2677  
 Db 6877 CTA-ACGAGACATCGGCGGCTCTGTGACAGACAGATGCTCGGGCAAGATCAGGCC 6935  
 QY 2678 SerYrTrIleProPheLeuLeuArgGluGlnuSerAlaLeuAsnProGly 2693  
 Db 6936 AGCTACATCCCTTCTTGTGCTGAGGAGAGTCCGCACTGAACCTGGG 6983

RESULT 6  
 ABLK15177 standard; DNA; 9321 BP.  
 ID ABLK15177 standard; DNA; 9321 BP.  
 XX ABLK15177:  
 AC  
 DE Human REPTR 9 cDNA sequence.  
 XX  
 XX REPTR; human; anti-inflammatory; cyrostatic; immunosuppressive;  
 KW antiviral; anti-HIV; antiarthritis; anticonvulsant; nootropic;  
 KW neuroprotective; anti-allergic; antibody; immunogen; endometriosis;  
 KW gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KW Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KW endocrine disorder; hypochalasia disorder; Kallman's disease;  
 KW autoimmune disease; inflammatory disease; infertility; receptor;  
 KW acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KW allergy; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KW systemic lupus erythematosus; cell proliferative disorder;  
 KW cancer; developmental disorder; Duchenne muscular dystrophy; gene;  
 KW Becker muscular dystrophy; neurological disorder; epilepsy;

KM		Alzheimer's disease; Huntington's disease; reproductive disorder; ss.
OS	Homo sapiens.	
XH	Key	Location/Qualifiers
FH	CDS	233..9033 /*tag= a /product= "REPTP9 protein"
FT	Pt CDS	
XX	MW0200198354-A2.	
PN		
PD	27-DEC-2001.	
XX		
PF	21-JUN-2001; 2001WO-US19942.	
XX		
PR	21-JUN-2000; 2000US-214027P. 25-AUG-2000; 2000US-228045P. 12-DEC-2000; 2000US-255104P.	
PR	(INCY-) INCYTE GENOMICS INC.	
PA		
PI	Griffen JA, Kallikch CM, Tribouley CM, Yue H, Nguyen DB, Tang YT, Lal P, Policky JL, Azimati Y, Lu DM, Grazi R, Igo MG, Butford N; Hafalia AJA, Baughn ME, Bandman O, Patterson C, Yang J, Xu Y, Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Duggan BM, Lu Y;	
PI	WP1; 2002-090432/12.	
DR	P-PSDB; AAU74826.	
PT	The diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell proliferative (e.g. cancer disorders -	
XX	Claim 65; Page 151-154; 157pp; English.	
CC	This invention relates to twelve human receptors cDNA sequences referred to as REPTR-1 to REPTR-12, and the proteins encoded thereby. The proteins of the invention may have anti-inflammatory, cytostatic, immunosuppressive, antiviral, anti-HIV, antirheumatic, muscular active general, anticoagulant, nociceptive, neuroprotective, antiallergic activities. The sequences of the invention may be used to produce REPTR agonists or antagonists, and the protein sequences may be used to raise anti-REPTR antibodies. These molecules and the REPTR polynucleotides and peptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma, Crohn's disease, irritable bowel syndrome, ulcerative colitis), endocrine (e.g. hypothyroidism disorder, Kallmann's disease), autoimmune/inflammatory (e.g. acquired immune deficiency syndrome (AIDS)), rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus, multiple sclerosis, systemic lupus erythematosus), cell proliferative (e.g. cancer), developmental (e.g. Duchenne and Becker muscular dystrophy), neurological (e.g. epilepsy, Alzheimer's disease, Huntington's disease) and reproductive (e.g. infertility, endometriosis) disorders. Numerous other examples of each disorder are given in the specification. The present sequence represents the human REPTR9 cDNA sequence of the invention.	
XX	SQ Sequence 9321 BP; 1780 A; 3008 C; 2776 G; 1757 T; 0 other;	
Alignment Scores:		
Pred. No.:	0 Length: 9321	
Score:	1958.00 Matches: 2886	
Percent Similarity:	96.91% Conservative: 0	
Best Local Similarity:	96.91% Mismatch: 17	
Query Match:	66.99% Indels: 92	
DB:	24 Gaps: 0	
US-09-916-849A-3 (1-2923) x ABKLS177 (1-9321)		
OY	13 ProProProProleuleuLeuLeuLeuLeuLeuLeuLeuProProProleuLength: 32	
Dh	262 CGCGCGCGCGCCTGCATGTGGTCAGTCACTGCCCGCACCATATTGGGAAC 32	

QY	33	GIATVALGYPProCysAsgSerLeuG1SerArgL1YArgG1YSerSerG1YALAaCYALA	52
Db	322	CAATGGGGGGCCCTTCGTTCCCTTGGGGTCCAGGGGAGAGAGGCTTCGCGGGGGCCTCGGCC	381
QY	53	PrometG1YTrpLeuCySProSerSerAlaSerAsnLeuTrpLeuYrTrhSerArgCys	72
Db	382	CCCATGGGGCTGGCTCTGTCCATCTCCAGCGCTGAACTCTGGCTCTACACAGCGCGTGC	441
QY	73	ArgAspAlaG1YTrhG1YLeuThG1YH1eLeuVal1ProH1sH1sAspG1YLeuArgVal	92
Db	442	AGGAGATGGGGCACTGAGCTGAGCTGGGCACTGTATCCCCAACAGATGGCTGAGGGTT	501
QY	93	TrpCySProG1YSerG1YAlaH1s1LeuProLeuProAlaProG1YG1YCySProTrp	112
Db	502	TGGTGTCCAGAAATCCAGAGGCCATATATCCCTTACACAGGTCTCTGAAGGGCTGGCCCTGG	561
QY	113	SerCySPArgLeuLeuG1Y11eG1YG1YH1sLeuSerProG1YG1YV1eLeuThLeuPro	132
Db	562	AGCGTGGCGCTCCCGGGGCAATGGAGGCACTTTCCCCAACAGGGCAAGCTCACACTGCC	621
QY	133	GIuG1YH1sPProCys1eLeuYsAlaPProArgLeuArgCysG1YSerCySLYsLeuAlaG1N	152
Db	622	GAGAGACACCCGTCCTTAAGGCTCCACGGCTCAGATGCCAGTCTCTGACAGCTGGCACAG	681
QY	153	AlaProG1YLeuArgAlaG1YG1YArgSerProG1YG1YSerLeuG1YG1YArgArgLYs	172
Db	682	GCCCCCGGGCTCAAGGGCAGGGGAAAGGCTCACAAAGAGTCCCTGGGGTGGCGCTCGGAAA	741
QY	173	ArgAsnVal1AsnTrpAlaProG1NPhG1NPhProSerTrpG1NAlaTrhValProG1Y	192
Db	742	AGGAATGAATTAACAGCCCCCACTTCCAGCCCCCAAGCTACACAGGCCACAGTCCGGAG	801
QY	193	AsnG1NProAlaG1YTrhProValAlaSerLeuArgAla11eAspProAspG1YG1Y	212
Db	802	AAACAGCAGAGGCAACCCCTGTTGCATCCCTGAGGGGCATCCAGCCGGAGAGGGTGAG	861
QY	213	AlaG1YArgLeuG1YrTrhMetAspAlaLeuPheAspSerArgSerArgG1NPhPhe	232
Db	862	GCAAGTGCAGCTGAGAGTACACCATGTGAAGCCCTCTTGAATAGCCGCTCCAAACAATCTTC	921
QY	233	SerLeuAspProVal1ThnG1YALAaVal1ThrTrhArg1eG1YLeuAspArgG1YTrhLYs	252
Db	922	TCCCTGGACCCAGTCACTGTGGCGAGTAAACACAGCCGAGAGGCTGGATCTGGAGACCAAG	981
QY	253	SerTrh1eVal1PheArgVal1ThrAlaG1NAspH1sG1YMetProArgSerAlaLeu	272
Db	982	AGCAACCAACGCTCTCAAGGGTCAACGGCCACAGGACCAACGGCATGCCCAACGAAAGCCCTG	1041
QY	273	AlaThrLeuThr11eLeuVal1ThrAspTrhAsnAspH1sAspProVal1PheG1YG1N	292
Db	1042	GCTACACTCACATCTTGATTACTGACACCAATACATAGACCTTGATGTTCCAGACAGAG	1101
QY	293	GIuYr1YsG1YSerLeuArgG1YAsnLeuG1YValG1YrTrG1YVal1LeuThValArg	312
Db	1102	GAGTACAAGAGAGAGCTCAAGGAGAACCTGAGGGTTGGCTATGAGGGTGCACACTGCAGG	1161
QY	313	AlaThrAspG1YAspAlaProProAsnAlaAsn11eLeuYrArgLeuLeuG1YSer	332
Db	1162	GCCACGAGATGGTAGGCCCTCCCAAGGCCAAATATTTGTAACGGCTGCTGGAGAGGGTCT	1221
QY	333	G1YG1YSerProSerG1YVal1PheG1Y11eAspProArgSerG1YVal11eArgTrhArg	352
Db	1222	GGGGGCGAGCCCTCTGAAGTCTTTGAGATGAGCCCTGCTCTGGGGGAGATCCGAACCGT	1281
QY	353	G1YProVal1AspArgG1YG1YVal1G1YSerTrG1YLeuThrVal1G1YAlaSerArgG1N	372
Db	1282	GGCCCTGTGATCGGGAAGAGGTGAATCTTACACAGCTGACGGTGAAGGCAAGTGAACAG	1341
QY	373	G1YArgAspProG1YProArgSerTrhThAlaAlaVal1PheLeuSerVal1G1YAspAsp	392
Db	1342	GGTGGGAGACCGGGTCTTCGGAATGACACAGCGGCTATTTTCCTTTCTGTGGAGATGAC	1401

Oy 393 AsnAspAsnAlaProGlnPheSerGluValArgGlyValValGlnValArgGluAspVal 412  
 Db 1402 AATGATTAATGCCCCAGTTTATGTAGAGAGCCCTATGTGTCTCAGATGAGGAGATGTG 1461  
 Oy 413 ThrProGlnValaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAla 432  
 Db 1462 ACTCCAGAGGGGCCAGTACTCCGAGTCAAGCCCTCGAATCGAGACAGAGGAGGACATGCG 1521  
 Oy 433 ValValHisGlySerTleuSerGlyAsnAlaArgGlyGlnPheGlyLeuAspAlaGln 452  
 Db 1522 GTGGTGCACTATAGCATCATGATGTGGCAATGCTCGGGGACAGTTTATCTGTGATCCCA 1581  
 Oy 453 ThrGlyAlaLeuAspValValSerProLeuAspGlyGlyLeuThrThyGlyLeu 472  
 Db 1582 ACTGAGGCTGTGAGTGTGTGAGCCCTCTTGACTATGAGACAGACAGAGAGATCAACCTTA 1641  
 Oy 473 ArgValaArgAlaGlnAspGlyGlyValArgProLeuSerAsnValSerGlyLeuValThr 492  
 Db 1642 CCGGTGCGAGACAGAGATGTGTGGCCCTCCCACTCTTATGTCTGTGGCTGTGAGCA 1701  
 Oy 493 ValGlnValLeuAspGlyAsnAspAsnAlaProIlePheValSerThrProPheGlnAla 512  
 Db 1702 GTACAGGCTCTGTGATATCAAGACAAATGCCCATCTTGTCAGACACCCCTTCCAGGCT 1761  
 Oy 513 ThrValLeuGlnSerValProLeuGlyGlyLeuValLeuHisValGlnAlaIleAspAla 532  
 Db 1762 ACTGTCCGTGAGAGCGTCCCTTAGGCTACCTGTCTTCATGTCTCAGGCTATCCAGCCT 1821  
 Oy 533 AspAlaGlyAspAspAsnAlaArgLeuGlnValArgLeuAlaGlyValGlyHisAspPhePro 552  
 Db 1822 GATGCTGTGACATATCCCGCCCTGAGATACCCCTTGCTGCGGGGTGGACATGACTTCCC 1881  
 Oy 553 PheThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGlyGlu 572  
 Db 1882 TTCACCATCAACAATGAGCAGCGCTGAGTCTGTGGCTGTGAGCTGAGCCGGAGGAA 1941  
 Oy 573 ValAspPheTyrSerPheGlyValGlnValArgAspHisGlyThrProAlaLeuThrAla 592  
 Db 1942 GTGATTTCTCAAGCTTGTGGGTAGAACCTCGAGCCATGAGCACTCCGACCTCACTCC 2001  
 Oy 593 SerAlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGln 612  
 Db 2002 TCCGCGCATGTCAAGGTACTCTCTGATGTCAACGACACATTCACCTTATCCAA 2061  
 Oy 613 ProGlnTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrVal 632  
 Db 2062 CCAAGGTACACAGTCCGCTCAATGAGATGAGCTGTGGGACAGAGTGTGTGACGCTG 2121  
 Oy 633 SerAlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThr 652  
 Db 2122 TCAAGCTGTGACCGGTATGCTCATGTCTCATCACTTACCGATCAACAGTGTGCAATACT 2181  
 Oy 653 ArgAsnArgPheSerIleThrSerGlnSerGlyGlyValLeuValSerLeuAlaLeuPro 672  
 Db 2182 CGAAACCCCTTCTCATCAACAGCCAAAGTGTGTGGCTGTGATCCCTTGCCTGTGCA 2241  
 Oy 673 LeuAspTyrGlyLeuGlnArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArg 692  
 Db 2242 CTGAGCTCAAACTTGAGCGGAGATGTGTGTGGCTGTATCCGCTCGAGTGGCACTCGG 2301  
 Oy 693 GlnAspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPhe 712  
 Db 2302 CAGGACAGGACAGATGTGTGTGATGTCAACCGACCGCAACCCATCGTCTGTCTTT 2361  
 Oy 713 GlnSerSerHisGlyThrThrValAsnValAsnGluAspArgProAlaGlyThrThrValVal 732  
 Db 2362 CAGAGCTCCCATATACAGTGAATTTATATGAGACCGGCGGACGACCAACGAGTGTG 2421  
 Oy 733 LeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGlu 752  
 Db 2422 CTGATCAACGCGCACGATGAGGACACAGGTGGAATGCCGATCACTTCAATGAG 2481  
 Oy 753 AspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlu 772

Db 2482 GACAGCATCCCCAGTTCGCGATCGATGACAGACAGGGGGCTGTACACCAAGGCTGAG 2541  
 Oy 773 LeuAspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyLe 792  
 Db 2542 CTGAGCTACGAAAGACCAAGTCTTACACCTGTGCATTAATCTGTGCGGAACTAGGCAAT 2601  
 Oy 793 ProGlnTyrSerAspThrThrTyrLeuGlnIleLeuValAsnAspValAspAspAsnAla 812  
 Db 2602 CCCAGAAAGTCCGACACCACTTACTTGAGATCTCTGTGAGACAGTGAATGACATGCC 2661  
 Oy 813 ProGlnPheLeuAspAspSerTyrGlnGlySerValTyrGluAspValProPheThr 832  
 Db 2662 CCTCAGTTCCTCGACGACTCTTACCAAGGCGAGTGTCTATGAGATGTGCCACCTTCACT 2721  
 Oy 833 SerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheThr 852  
 Db 2722 AGGCTCTGAGATCTGACCACTGATGTGATTTCTGACTTAATGGCGAGGCTTCTTAC 2781  
 Oy 853 ThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGlnSerThrSerGlyIle 872  
 Db 2782 ACCTTCCAGAGAGCGACAGATGAGAGCGGTGACTTTATGTGTGATCCACTCAGGATC 2841  
 Oy 873 ValArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyr 892  
 Db 2842 GTGCGAAGCTTACGAGAGCTGATCGAGAGAACGTGGCCGATGTCTTGTGCGGGCATAT 2901  
 Oy 893 AlaValAspArgGlyMetProProAlaArgThrPheMetGluValThrValThrValLeu 912  
 Db 2902 GCAGTGTGACAAAGGAGAGTCCCCAGCCCGCACCTATGAAAGTACATCATGTGTG 2961  
 Oy 913 AspValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGlnGlu 932  
 Db 2962 GATGTGATGACATATCCCTCGTCTTTGAGAGAGATGATTTGATGTGTGTGTGGAAG 3021  
 Oy 933 AsnSerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlyGlyThr 952  
 Db 3022 AACAGCCCAATGGGTGAGCGGTGCGGGGTGCACACCTGACCCCGATGAAGGACCC 3081  
 Oy 953 AsnAlaGlnIleMetTyrGlnIleValGlnGlyAsnIleProGluValPheGlnLeuAsp 972  
 Db 3082 AATGCCAGATTAATGTACAGATTTGTGAGAGGACAAATCCCTGAGTCTTCCACTGTGAC 3141  
 Oy 973 IlePheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyr 992  
 Db 3142 ATCTTCCCGGGAGGTGACAGCGCTGTGTGACTTATCACTACAGACCGGCTGAGTAC 3201  
 Oy 993 ValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArg 1012  
 Db 3202 GTCTGTGTCAATCCAGGCGACGTGAGCTCTGTGTGAGCGGGCTTACAGTCCAGTCCGC 3261  
 Oy 1013 LeuLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGlnIleLeuPheAsnAsn 1032  
 Db 3262 CTCTTTCACCGCAATGACAAACCAACAGTGTGGGCAACTTGTGATCTTTTCAACAAAC 3321  
 Oy 1033 TyrValThrAsnArgSerSerSerPheProGlyGlyValaIleGlyArgValProAlaHis 1052  
 Db 3322 TATGTCAACATGCTCAACAGAGTCTCTGTGGGTGCCATTTGGCCAGATACCTGCCCAT 3381  
 Oy 1053 AspProAspIleSerAspSerLeuThrTyrSerPheGlnArgGlyAsnGluLeuSerLeu 1072  
 Db 3382 GACCTATATCTCAGATGATGTGACTTACAGCTTTTGAAGGGGGAATGAATCAAGCTCG 3441  
 Oy 1073 ValLeuLeuAsnAlaSerThrGlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArg 1092  
 Db 3442 GTCTGTCTCAATGCTTCAACGGGTGAGCTTAAGCTTAAGCCGCACTGAGACAAACCGG 3501  
 Oy 1093 ProLeuGlnAlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGln 1112  
 Db 3502 CCTGTGAGGCGCATCATGACGATGTGTGTGACAGCGCGTACACAGCGTGAACGGCCAG 3561  
 Oy 1113 CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArg 1132



D 3562 TGGCGCTGCGTGTGACCATCATCATGAGTAGTGTCTCCACCAAGCATCAGCGTGGC 3621  
Q LeuGluAspMetSerProGluArgPheLeuSerProLeuGluGlyLeuPheIleGlnAla 1152  
D 3622 CTGGAGGACATGTACCCGAGCGCTCTGTGCACACATGCTGAGGCTCTTCACTCCAGGCG 3681  
Q ValAlaAlaThrLeuAlaThrProProAspHisValAlaValPheAsnValGlnArgAsp 1172  
D 3682 GTGGCGCGCACGCTGGCGCACCGGACCAAGCTGTGTGTCTTCAACGTACAGCGGGAC 3741  
Q ThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProGly 1192  
D 3742 ACCGACCGCCCGCGGGCGCATCTCTCAAGCTGAGCTGTGCGGGCGCGCGCGCGG 3801  
Q ProGlyGlyProPheLeuProSerGluAspLeuGlnGlyLeuGlyLeuAsn 1212  
D 3802 CCGGGGGCGGGCGCGCTTCTGCTCTGAGGACCTGCGAGGCGCGCTTATCTCAAC 3861  
Q ArgSerLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsnIleCys 1232  
D 3862 CGGAGCTGTGACGGCATCTGGGACAGGGGCTGTGCCCTTCCAGACATCTGC 3921  
Q LeuArgGluProCysGluAsnIleMetArgCysValSerValLeuArgPheAspSer 1252  
D 3922 CTGCGGAGCGCTGCGAAGCTACATGCGCTGTGCGTGTGCGCTTCCGACTCTCC 3981  
Q AlaProPheIleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeu 1272  
D 3982 GCGGCTTCACTGCTCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4041  
Q ArgCysArgCysProProGlyPheThrGlyAspIleCysGluThrGluValAspLeuCys 1292  
D 4042 CGCTGCGCGTGGCGCGCGCGCTTCAAGGGTGACTACGCGGACCGAGGTGGACCTTGC 4101  
Q TyrSerArgProCysGlyProHisGlyArgCysArgSerArgGlyGlyGlyTyrThrCys 1312  
D 4102 TACTCGGCGCGCTGTGGCGCGCGCGCTGCGCGCGCGCGCGCGCTCACTGC 4161  
Q LeuCysArgAspGlyTyrThrGlyGluHisCysGlyValSerAlaArgSerGlyArgCys 1332  
D 4162 CTCTGCTGTATGCTACCGGCTGAGCATCTGTAGGTGAGTGTCTGCTCAAGCGCTTGC 4221  
Q ThrProGlyValCysIleAsnGlyGlyTyrCysValAsnLeuValGlyGlyPheLeu 1352  
D 4222 ACCCGCGGTGTGCAAGAAATGGGGCAGCTGTGTCAACCTGCGTGGCGGTTCAAG 4281  
Q CysAspCysProSerArgIleAspPheGluIleProCysArgGlnValThrThrArgSerPhe 1372  
D 4282 TGCATTTGCCATCTGGAGACTTCGAGAGCGCTTACCTGCGAGTGAACAGCGCGACTTC 4341  
Q ProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAla 1392  
D 4342 CCGCGCCACTCTTCACTCACTTTCGCGGCTGTGGCGAGCGTTCACCTTCACTTGC 4401  
Q LeuSerPheAlaThrIleGlyIleArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGly 1412  
D 4402 CTCTCGTGTGCAAAAGAGCGCGAGCGGTTGCTGTGTACATGTGGCGTTCATAGAG 4461  
Q LysHisAspPheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPheSerAla 1432  
D 4462 AAGCATACCTTGTGGCGCTGTGAGGTATCCAGGAGCGGTCCAGCTCACTTCTGCA 4521  
Q GlyGluSerThrThrThrValSerProPheValProGlyGlyValSerAspGlyGlnTyr 1452  
D 4522 GGGGAGTCAACCAACAGGTGTCCCATTTGCTGCGGAGGAGTCAATGATGGCAATGG 4581  
Q HisThrValGlnLeuGlyTyrTyrAsnIleProLeuLeuGlyGlnThrGlyLeuProGln 1472  
D 4582 CATACGCTGACGCTGAATATCAATAGCACTGTGTGCTCAAGAGGCTCCCAAG 4641  
Q GlyProSerGluGlnIleValAlaValThrValAspGlyCysAspThrGlyValAla 1492  
D 4642 GGGCCATCAAGCAAGAGGTGTGTGTGTACCTGTGTGTGTGTGTGTGTGTGTGTGT 4701

Q 1493 LeuArgPheGlySerValLeuGlyAsnIleTyrSerCysAlaGlnGlyThrGlnGlyGly 1512  
D 4702 TTGGCTTGGATCTGTCTGGGCACTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 4761  
Q SerIleValSerLeuAspLeuThrGlyProLeuLeuLeuGlyGlyValProAspLeuPro 1532  
D 4762 AGCAAGAACTCTGTGATCTGACGGGGCGCTGTACTAGGGGGGTGTGCTGACCTGCC 4821  
Q GlnSerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAsp 1552  
D 4822 GAGAGCTTCCAGTCCGAATGGCGGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4881  
Q SerArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyTyrThrValProGlyCysPro 1572  
D 4882 AGCGGCACTAGACATGCTGACTGACTTATTCGACAAATGGCAACCTGTGCTGTCCCT 4941  
Q AlaIleValAsnValCysAspSerAsnIleCysHisAsnGlyGlyTyrCysValAsnGln 1592  
D 4942 GCCAAGAAAGATGTGTGTGACAGCAACCTTGCACAAATGGGGGCACTTGGCTGAACAG 5001  
Q TyrAspAlaPheSerCysGlyCysProLeuGlyPheGlyGlyIleValSerCysAlaGlnGly 1612  
D 5002 TGGAGCGCTTCACTGACGAGATGCCCCCTGTGGCGGCGAAGACTGCGCCAGAA 5061  
Q MetAlaAsnProGlnHisPheLeuGlySerSerLeuValAlaTyrHisGlyLeuSerLeu 1632  
D 5062 ATGGCAATCCAGACACTTCTTGGGACAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 5121  
Q ProIleSerGlnProTyrPheLeuSerLeuMetPheArgThrArgGlnAlaAspGlyVal 1652  
D 5122 CCCATCTCCCAACCTGTGATCTCAAGCTCAATGTTCCGACGCGCGCAAGCGCTGTGC 5181  
Q LeuLeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHis 1672  
D 5182 CTGTGCAAGCCATCCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5241  
Q ValMetLeuSerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGly 1692  
D 5242 GTGATCTGATGCTGAGGAGGACAGGGGCTTCAAGGCTTCTCTCTCTCTCTCTCTCT 5301  
Q ArgAlaAsnAspGlyAspTyrHisHisAlaGlnLeuAlaLeuGlyValAspGlyTyrPro 1712  
D 5302 CGGGCCAAATGACGCTGTACCTGACCAATGCAAGCTGTGCACTGGAGCGCGGGCGCT 5361  
Q GlyHisAlaIleLeuSerPheAspTyrGlyGlnArgAlaGlnGlyValAsnLeuGlyPro 1732  
D 5362 GGCATGCAATCTGTCTCTTCAATTAAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5421  
D 5422 CCGCTCAATGCTGTGACCTGAGCAACATTAACATTAAGTGGGGAATATCTGGGCGAGCG 5481  
Q GlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlu 1772  
D 5482 GGTGTGCGCGT 5541  
Q GlyValAsnSerLeuAspProSerHisGlyGlnSerIleAsnValGluGlnGlyCysSer 1792  
D 5542 GGGGTTTAACTGTGATCTCCAGCCATGGGAGAGCATCAAGTGAAGCAAGGCTGTAGC 5601  
Q LeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTyr 1812  
D 5602 CTGCTGACCTTGTGTATCAAAACCGGTCTCTGTCAAGCTATGCGAGCAACGACTGG 5661  
Q AspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAspAsnValSerHisValCys 1832  
D 5662 GACAGCTATTCTGACGCTGTATCAAGGTATCTATAGTGAACAATCTATATGTGTGT 5721  
Q AspLeuAsnProCysGlyGlnSerValCysThrArgIleAspProSerAlaProHisGly 1852  
D 5722 GACTGAACCTGTGTAGGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5781



QY	1853	YTRTHRCYGGIUCYSPROProAsnTYrLeuGIYPRCYGGIUTHArgILeaSerGI	1872
Db	5782	TATACCTCGAGAGTGTCCCAAAATTACTTGGGCGATCTGGAGACCAAGATTGACAG	5841
QY	1873	ProCSPROArgSGIYTPRTProGIYHisProThRCysGIVProCYSAsnCYAspValSer	1892
Db	5842	CCTTGTCCCCGTGGCTGGTGGGGAACATCCCAATGTGGCCCAATGCACATCTGAATGCACG	5901
QY	1893	LYSGIYpHeaSPProAspCYAsnLYSThSerGIYGIUCYHisCYeLYSGIYAsnHis	1912
Db	5902	AAAGGCTTTGACCCAGACTGCACAAAGCAAGCGCGAGGTGCATCTGAAAGAACAC	5961
QY	1913	TYTAAGProProGIYSerProThRCYSLeuLeuCYAspCYeTYRProThGIYSerLeu	1932
Db	5962	TACCGGCCCCCAGGAGCGCCCACTGCCTCTGTGTGACTGTACCCACAGGCTCTTGG	6021
QY	1933	SerATGValCYAspProGIYAsnPLYGIUCYSProCYeLYSPROGIYValIIIGIYArg	1952
Db	6022	TCCAAGTCTGTGACCTCGAGGATGGCCAGTGTCCATGCACAGCCAGGTCATGGGGCT	6081
QY	1953	GIUCYAspATGACYAspAsnProPheAlaGIUValIThThrAsnGIYCySGIU-----	1970
Db	6082	CAGTGTGACCGGTGTGACACACCTTTGTCTAGAGTCAACACCAATGCTGTGAAGGGGCC	6141
QY	1971	--ValAsnTYrAspSer--CYSPROArgAlaIle-Glu--AlaGIYIleTRTPProAr	1988
Db	6142	TYGTTTG--CT--AG--TT--ACTGTCCCG--GCCCA--TGAAGTGTGGCC--TC--CA--GC--AG	6190
QY	1988	GFTr-----Arg---PheGIY-----LeuProAla-AlaAlaProCYSPR	2000
Db	6191	AACTCTCAGCCAGTCTCAGGGGGCTTCTT--GTGTGTCTCCCTGA--GGCGCG--CCCTT--TT	6246
QY	2000	OLYSGIYSerPhe-----GIYThrAlaValArgHisCYAspGIUHisArgGIYTRPLeu	2018
Db	6247	--GG--CT--TCTTTCCCCCAGGAGACTGTCTGTGGCCACTGTGATGAGACACAGGGGGTGGCT	6303
QY	2019	ProProAsnLeuPheAsnCYeThrSerIleThrPheSerGIUeLYSGIYPheAlaGIU	2038
Db	6304	CCCCCAAACTCTTCACTCACTGACGTCACATCACTTCTCAGAACTGAAGGGCTTGGCTGAG	6363
QY	2039	ArgLeuGIUArgAsnGIUSerGIYLeuAspSerGIYArgSerGIUAsnAlaIleuLeu	2058
Db	6364	CGGCTTACAGCGGATGAGTCAAGGCTTCACTCACTCAGGGCGCTCCAGCAAGCTTGTCTCTC	6423
QY	2059	LeuArgAsnAlaThrGIUHisThrAlaGIYTYRPhGIYSerAspValIYValAlaTYR	2078
Db	6424	CTGCCCAACGCCACGCCACACACAGCTGGCTACTTCCGACACGACGTCAAGGTGGCTTAC	6483
QY	2079	GIUeLYAlaThrArgLeuLeuAlaHisGIUSerThrGIUArgGIYPheGIYLeuSerAla	2098
Db	6484	CAGCTGGCCACGGCGGCTGTGGCCACGAGAGCACCCAGCGGGCTTTGGGCTGTCTGCC	6543
QY	2099	ThrGIUAspValHisPheThrGIUAsnLeuLeuArgValGIYSerAlaLeuLeuAspThr	2118
Db	6544	ACACAGGACGGACCTTCACTGAGATTYGTGTGGGGGGGACACGGCGCTCTCTGACACA	6603
QY	2119	AlaAsnLYSTArgHisSTRGIUeLYIleGIUInThrGIUArgIYValIATRPLeuLeu	2138
Db	6604	GCCAAACAGCGGCACTGGAGACTGTATCCAGACACAGAGGGGTGCACCGCTGGCTGTCTC	6663
QY	2139	GIUHisTYRGIUAlaTYRAlaSerAlaLeuAlaGIUAsnMetArgHisThrTYRLeuSer	2158
Db	6664	CAGACCTATGAGGCTTACGCCAGTGGCTTGGCCCAAGCAATGGGGACACCTACCTAAGC	6723
QY	2159	ProPheThrIleValIThrProAsnIleValIIISerValValArgLeuAspLYSGIYAsn	2178
Db	6724	CCCTTCAACATCGTCACGCCCAAAATGTCACTTCCGTAAGTGCCTTGGACAAAGGGACAC	6783
QY	2179	PheAlaGIYAlaLYSLeuProArgTYRGIUAlaLeuArgGIYGIUArgIYProProAspLeu	2198
Db	6784	TTTGTCTGGGGGCAAGCTTCCCCCGCTTGAAGGCCCTTGGTGGGAGACAGCCCCGGACCTT	6843
QY	2199	GIUThrThrValIleLeuProGIUSerValPheArgGIUThrProProValIArgPro	2218

Db	6844	GAGACAACAGTCATTCTGCGCTGAGACTGTCTTCAGAGAGAGGCCCGCTGGTCAAGGCC	6903
Qy	2219	AlaGlyProGlyGlyAlaGlnGlyProGlyGlyLeuAlaArgArgGlyArgArgHisPro	2238
Db	6904	GCAAGGCCCCGGAGAGGCCCAAGAACCAAGAGAGCTGGCAAGGCCAACAAGCAAGGCCACCG	6963
Qy	2239	GluLeuSerGlnGlyGlyAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeu	2258
Db	6964	GAGCTGAAGCCAGGGGTGAGGCTGTGGCCAGCGTCATCATTCACGCCACTGTGGCGGGCTA	7023
Qy	2259	LeuProHisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIle	2278
Db	7024	CTGCTCACTAACTATGACCTTCGACAAAGGCGACTTGAGAGTCCCAACGCCCATCATC	7083
Qy	2279	AsnThrProValValSerIleSerValHisAspAspGluLeuLeuProArgAlaLeu	2298
Db	7084	AACACACCCCGTGGTGAAGCATCAGCGCTCATATATATAGAGAGCTTCGCCCCGGAGCCCTG	7143
Qy	2299	AspLysProValThrValGlnPheArgLeuLeuGluThrGlnArgThrLysProIle	2318
Db	7144	GACAAACCCGTCAAGGTGACAGTTCCGCTGTGGAGACAGAGAGCGGACCAAGCCATC	7203
Qy	2319	CysValPheTyrPasnHisSerIleLeuValSerGlyThrGlyTyrPserAlaArgGly	2338
Db	7204	TGTGTCTTCTGAACCAATTCATCTGTGTCACTGTCAGTGGCAAGGTGGCTGTGGCCAGAGGC	7263
Qy	2339	CysGluValValPheArgAsnGlySerHisValSerCysGlnCysAsnHisMetThrSer	2358
Db	7264	TGTGAGTGTGCTTCCGATAGAGAGCAAGTCAAGTGCACAGTGCACCAATGACAGC	7323
Qy	2359	PheAlaValLeuMetAspValSerArgGlyLys-----	2369
Db	7324	TTCCGTGTGCTCAAGAGAGTTTCTCCGCGGAGAGTCCGGCCACAGGGGACGTCGACAG	7383
Qy	2370	-----AsnGlyGluIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThr	2387
Db	7384	CCGTGGAATGGGAGATCCCTGCGCACTGAAACACTGACATACGGTCTTACGGTGTACCC	7443
Qy	2388	LeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuAlaGlySerAsn	2407
Db	7444	TTGGTGCCCTTCGTCTCACTCTTCTTCTCTCACTCTCTTGGGTATCTGCGCTCCAC	7503
Qy	2408	GlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu	2427
Db	7504	CAACACGGCAATCCGACCTTAACCTGACAGCTGCCCTGGGCTCAGCTGCTCTTCTC	7563
Qy	2428	LeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis	2447
Db	7564	CTGGGAATTCACACGAGCTGACCTCCCTTTGGCTGCGACAGTCAATGGCATCTGCTGCAC	7623
Qy	2448	PheLeuTyrLeuCysThrPheSerTyrPheAlaLeuLeuGlnAlaLeuHisIleLeuTyrArgAla	2467
Db	7624	TTCTGTGTAACCTGTGACCTTTCTCTGGGCTCTCTGAGAGCCCTTGACCTGTACCCGGGCA	7683
Qy	2468	LeuThrGlnValArgAspValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyr	2487
Db	7684	CTCACGTAGAGTGGCGCAATGCAACACGGGCCCAATGCGCTTCTAATCAATGCTGGGCTGG	7743
Qy	2488	GlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGlnGlyTyrGlyAsn	2507
Db	7744	GGCGTGCTGCTTCATCAACAGGAGCTAGCGTGGGCTGGACCCCGAGGGCTACGGGAAAC	7803
Qy	2508	ProAspPheCysTyrPheLeuSerIleTyrAspThrLeuLeuTyrPserPheAlaGlyProVal	2527
Db	7804	CTTACATCTGTGGGTGCTCTCACTATGACACCCCTCATCTGAGAGTTTGTCTGGCCGGGTG	7863
Qy	2528	AlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla	2547
Db	7864	GCTTTTGCCGTCTTCATGATGTGCTTCTCTAATCTCTGGCGGCCCGGGGCTCTGTGCT	7923
Qy	2548	AlaGlnArgGlnGlyPheGluValGlyGlyProValSerGlyLeuGlnProSerPheAla	2567

Db 7924 GCCCAGCGGAGGCTTTGAGAAAGATCTGCTCGGGCTTGAGCCCTCTTCGCC 7983  
 Qy 2568 ValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValaAsnSerAspThr 2587  
 Db 7984 GTCTCTCTGCTGAGCGGCGGAGGCTGCTGCTCTCTGTCACACGACGCC 8043  
 Qy 2588 LeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
 Db 8044 CTCCTCTTCACCTACCTTTGCTACCTGCAATTCAGAGGCCCTTCATCTTCCTC 8103  
 Qy 2608 SerTyrValValLeuSerTyrGluValArgValAlaLeuLeuAlaCysSerArgTyr 2627  
 Db 8104 TCTATGTTGTGTGTGTGACAGAGGTCCTCGAAGACCTCAAGCTTCTGACGCCCAAG 8163  
 Qy 2628 ProSerProAspProAlaLeuThrThrIleSerThrLeuThrSerTyrAsnCysPro 2647  
 Db 8164 CCAGCCCTGACCTGCTGTGACCAACCACTCACTGACCTGCTGCTCAACACGCCCC 8223  
 Qy 2648 SerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHis 2667  
 Db 8224 AGCCCTCAGCAGATGGGGCGGCTGTACAGCCCTACGGAGACTCGGGCGGCTCTGACAC 8283  
 Qy 2668 SerThrSerArgSerGlyTyrSerGlnProSerTyrTleProPheLeuLeuArgGlu 2687  
 Db 8284 AGCACAAGTGTCTCGGGCAGAGATCGAGCCAGCTACATCCCTTGTGTGAGGGAGAG 8343  
 Qy 2688 SerAlaLeuAspProGlyGlnGlyProProGlyTyrGlnGlyAspProGlySerLeuPheLeu 2707  
 Db 8344 TCCGCACTGACCTTGCCCAAGGAGGCCCCCTGCGCTGGGGAGATCCAGGCAAGCTGTTCCTG 8403  
 Qy 2708 GlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspSerLeuGln 2727  
 Db 8404 GAAGGTCAAGACCCAGCAGCATGATCTCGACAGGATCCGACAGTGACCTGTCTTAGAA 8463  
 Qy 2728 AspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlyGlnGlnGln 2747  
 Db 8464 GACGACCAAGAGTGTCTCTATGCTCTTACCTCACTCAACAGTACAGAGGAGAGAGAG 8523  
 Qy 2748 GlnGlnGlnGlnGlnAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGlyPro 2767  
 Db 8524 GAGGAGAGAGAGAGGCGCCCTTCTCTGAGAGACAGGCGTGGATGCTGCTGGGGCT 8583  
 Qy 2768 GlyAlaGlnArgLeuProLeuHisSerThrProLeuAspGlyGlyProGlyProGlyTyr 2787  
 Db 8584 GAGCAGAGAGAGCTGCCCCCTGCAAGTATCCCAAGATGGGGGCCCAAGGCTGCAAG 8643  
 Qy 2788 AlaProTyrProGlyAspPheGlyThrThrAlaIleGlnSerGlnGlnGlnGlnAlaPro 2807  
 Db 8644 GCCCCTGCGCAGAGAGCTTTGGGACCAAGCAAGAGTGTGGGCAACGGGGCCCT 8703  
 Qy 2808 GlnGlnArgLeuArgGlnArgGlnAlaLeuSerArgGlnGlySerLeuGlyProLeu 2827  
 Db 8704 GAGGAGCGGCTGCGGAGAGATGAGATGCTCTCTGAGAGGGGCTCCTAGGCCCTT 8763  
 Qy 2828 ProGlySerSerAlaGlnProHisIleGlyTyrIleLeuAlaIleCysLeuProThrIle 2847  
 Db 8764 CCGGCTCTCTTCTCCAGCCTCAAAAGGATCTTAAGAGAGTGTGCCCAACATC 8823  
 Qy 2848 SerGlnIleSerSerLeuLeuArgLeuProLeuGlnGlnCysThrIleSerSerArgGly 2867  
 Db 8824 AGGAGAGAGAGAGCTCTGCGGCTCCCTCTGAGAGAGCAAGAGGCTTCCCGAGG 8883  
 Qy 2868 SerSerAlaSerGlnGlySerArgGlyGlyProProProArgProProArgGlnSer 2887  
 Db 8884 TCTCCGCTGTGTAGGGCAGCGGGGGGCGCCCTCTCCGCCCAACCGCCCGGAGAGG 8943  
 Qy 2888 LeuGlnGlnGlnLeuAsnGlyValMetProIleAlaMetSerIleIleValaGlyThrVal 2907  
 Db 8944 CTCACAGAGAGCTGAACGGGGGCTATGCCATGCGCATGAGCATCAAGAGGACAGCGTG 9003  
 Qy 2908 AspGlnAspSerSerGlySerGln 2915  
 Db 9004 GATGAGAGCTCTCAGGCTCCGA 9027

RESULT 7  
 ID AACT6401 standard; cDNA, 9121 BP.  
 XX  
 AC AACT6401;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF1956 polynucleotide sequence SEQ ID NO:3911.  
 XX  
 KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;  
 KW vulerary; antipneumatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteoporotic; antidiabetic; immunosuppressant; cardiac;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antineuritic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000MO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CUDAGEN CORP.  
 XX  
 PI Shimketa RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR P-PSDB; AAB42192.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 3061-3067; 5507pp; English.  
 XX  
 CC AACT6446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytosolic; hepatotropic; vulerary;  
 CC antipneumatic; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteoporotic; anticonvulsant; antidiabetic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineuritic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.



Db	4507	TTTGGCTAAGCTCACCAACCAATGGCTGTGAATGAATTATGACAGCTGCCACGAGCAATT	4566
Qy	1981	GLUALAGLYLLETPTTPProAxyGThrArgPheGlyLeuProAlaAlaProCyAPro	2000
Db	4567	GAGGCTGGAGTCTGGTGGCCCCGTAACCGCTTCGGGCTGGCTGGCTGGCTCCGTGCC	4626
Qy	2001	LyseGlySerPheGlyThrAlaValArgHisCyAAspGlyLHisArgGlyLTPLeuProPro	2020
Db	4627	AAAGGCTCTTTGGGACGTGCTGTGGCCACTGTATGATGACAGAGGGGTGGCTCCCCCA	4686
Qy	2021	AsnLeuPheAsnCySerThrSerLethrPheSerGlyLeuLysGlyPheAlaGlyLysGlu	2040
Db	4687	AACTCTTCAACTGCACGTCCATCATCTTCCAGAACTGAAGGGCTTGCTGAGGGGCTA	4746
Qy	2041	GlnArgHisGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg	2060
Db	4747	CAGGGAAATGATGACAGGCTGTAAGCTACAGGGGGCTCCACGAGCTAAGCTCTCTGGGC	4806
Qy	2061	AsnAlaThrGlnHisGThrAlaGlyLYrPheGlySerAspValLysValAlaTYrGlyLeu	2080
Db	4807	AAAGCCAGGACGACACACAGCTGGCTACTTCGGGACGACGCTCAAGTGGCTTACAGCTG	4866
Qy	2081	AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln	2100
Db	4867	GCCACGGGGCTGGCTGGCCACGAGAAGCCACGAGGGGGCTTGGGGCTGTGCTCCACAG	4926
Qy	2101	AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
Db	4927	GACGTGCACTTCACTGAGAACTGTGCTGGGGTGGGACGGCTCTGTGAACAGCCAAAC	4986
Qy	2121	LysArgHisGTPGlyLeuLysGlnGlnThrGlnGlyGlyThrAlaATPLeuLeuGlnHis	2140
Db	4987	AAAGGACACTGGAGGTGATCCAGACAGCAGAGGGGTGCACCGCTGGCTCTCCAGAC	5046
Qy	2141	TYrGlnAlaTYrAlaSerAlaLeuAlaGlnAsnMetArgHisGThrTYrLeuSerProPhe	2160
Db	5047	TATAGAGGCTTACGGCAATGCCCCGTGGCCCAAGATCGGGCACCTTACTTAAGCCCCCTTC	5106
Qy	2161	ThrLysLeuAlaThrProAsnLysValLysSerValArgLeuAspLysGlyAsnPheAla	2180
Db	5107	ACACTCGTCAAGCCCAACTTGTCTACCTCCGTAATGTGGCCCTTGACAAAGGAACTTGGCT	5166
Qy	2181	GlyAlaLysLeuProArgTYrGlyAlaLeuArgGlyGlyGlnProProAspLeuGlyThr	2200
Db	5167	GGGGCCAAAGCTGCCCCGCTACGAGGGCCCTGTGGTGGGAGCAGCCCGCGAAGCTTGAGCA	5226
Qy	2201	ThrValLysLeuProGlnSerValPheArgGlyThrProProValAlaArgProAlaGly	2220
Db	5227	ACAGCTCACTTGGCTGAGTCTGTCTTCAAGAGAGAGCCCCCGTGTGTAGGCGCGAGGC	5286
Qy	2221	ProGlyGlyAlaGlnGlyProGlyGlyLeuAlaArgArgGlnArgArgHisProGlyLysLeu	2240
Db	5287	CCCGGAGAGGCCACGAGAGCCAGAGGAGGTGGCAGCGGACGCGACCCGGAAGCTG	5346
Qy	2241	SerGlnGlyGlyAlaValAlaSerValLysLysTYrArgThrLeuAlaGlyLeuLeuPro	2260
Db	5347	AGCCAGGGTGAAGGCTGTGGCCAGGCTCAATCTACCGCACCTGTGGCGGGCTACTGTGCT	5406
Qy	2261	HisAsnTYrAspProAspLysArgSerLeuArgValProLysArgProLysLysAsnThr	2280
Db	5407	CATAACTATGACCTTGACAAACGCGACGCTTGAGAGTCCCCCAAAAGCCGCAATCATCAACA	5466
Qy	2281	ProValValSerLysSerValHisAspAspGlyLeuLeuLeuProArgAlaLeuAspLys	2300
Db	5467	CCCGTGTGACACATCAAGGCTCAATGATATGAGGAGCTTTCGCCCCGGGGCCCTGGACAA	5526
Qy	2301	ProValThrValGlnPheArgLeuLeuGlnThrGlnGlyLysGlyThrLysProLysGlyVal	2320
Db	5527	CCCGTCAAGGTGCACTTCCGCTGTGTGAGACAGAGAGCGGACCAAGCCCAATCTGTGTCT	5586
Qy	2321	PheThrAsnHisSerLysLeuValSerGlyThrGlyGlyTTPSerAlaArgGlyCYAsGly	2340

Dd	5587	TTCTGGAAACATTGATTCCTGGTCAAGTGGCAACAGTGGCTGTGGCCAGAGGCTGTGAA	5464
Qy	2341	ValValPheArgAngIuSerHisValSerCysGlnCysAenHisMetThrSerPheAla	2360
Dd	5647	GTCGCTCTCCCAATGAGAGCCAGTCAAGCTGCCAGTGCACCAATCAACAAGCTTGCGCT	5706
Qy	2361	ValIleuMetArgPValSerArgArgGluAsnGlyValIuIleuProIleuValThrLeuThr	2380
Dd	5707	GTCCTCATGACGTTTTCTCGGCGGAGAAATGGGAGATCTTGCCACTGAAAGACTGACA	5766
Qy	2381	TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu	2400
Dd	5767	TACGGGCTTAGGTGATCACTTGCGTGGCCCTTGCTGACTCTTCTTCTCTCATCTCC	5826
Qy	2401	LeuArgGlyIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly	2420
Dd	5827	TTGGGTATCTCGGCTCCCAACACAGGGATCCGAGATCACTGACAGCTGGCTGGGC	5886
Qy	2421	LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaPheLeuProPheAlaCysThr	2440
Dd	5887	CTGGCTACGGCTGGCTTCTCTCTGGAGATCAACAAGCTGACTCTCCCTTTGGCTGGACA	5946
Qy	2441	ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGlu	2460
Dd	5947	GTCATTGGCACTCCGCTGCACCTTCCTGACTCTGCACCTTTTCTGGGCTCTGCTGAG	6006
Qy	2461	AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Dd	6007	GCTTGGCACTGTACCGGGCACTACTAGAGGTGGCCATGTACACACCGGCCCATGGCC	6066
Qy	2481	PheTyrTyrMetLeuGlyTyrTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Dd	6067	TTCTACTACATAGCTGGGCTGGGCGTGGCTCTTCACTACAGAGGCTAGCCGTGGGCTGG	6126
Qy	2501	AspProGluGlyTyrTrpGlyAsnProAspPheCysThrLeuSerIleTyrAspThrLeuIle	2520
Dd	6127	GACCCCGAGGAGCTACGGAAACCTTGACTCTGCTGGCTCTTCACATAGACAGCTCATC	6186
Qy	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu	2540
Dd	6187	TGGAGTTTGGCTGGCCCGGTGGCTTTGGCGGTCCGATGAGTGTCTTCTGTATCATCTCG	6246
Qy	2541	AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluValGlyProValSer	2560
Dd	6247	GCGGCGCGGGCTCTGTGCTGCCACAGGAGGGCTTGAAGAAAGGCTCTGTCTCG	6306
Qy	2561	GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuAlaLeu	2580
Dd	6307	GGCTTCAGGCTCTCTTCGCGCTCTCTGTGCTGAGCGGCACTGGCTCTGGCACTG	6366
Qy	2581	LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle	2600
Dd	6367	CTCTCTGTCAACACGACACCCCTCTCTTCCACTACACTTTGGCTACCTGCATTTGGCATC	6426
Qy	2601	GlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGlyValArgGlyValLeu	2620
Dd	6427	CAGGCGCCCTTCATCTTCTCTCTATAGTGTGTAGCAAGAGAGTCCGAGAAAGCACTC	6486
Qy	2621	LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu	2640
Dd	6487	AAGCTTGCTGCACCGGCAAGCCAGCCGCTGACACCTGTGTGACCAACAATGCACCCCTG	6546
Qy	2641	ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly	2660
Dd	6547	ACCTGCTCTACACTGCGCCAGCCCTTACGCAATGGCGGCTGTACACACCTCTACGGA	6606
Qy	2661	AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle	2680
Dd	6607	GACTGCGCGGCTCTCTGCACAGCAACAAGTGTGGGCAAGAATCAAGCCACGCTACATC	6666
Qy	2681	ProPheLeuLeuArgGluGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly	2700
Dd	6667	CCCTTCTTGTCTGAGAGAGAGTCCGCACTGAAACCTGGCCAAAGGCGCCCTGGGCTGGGG	6726

QY 2701 AspProGlySerLeuPheLeuGluGluGlnAspGlnGlnHisAspProAspThrAspSer 2720  
DB 6727 GATCCAGGAGGAGCTGTTCTCGAAGGTCAAGACACGACAGCATGATCTTCAACGAGACTCC 6786  
QY 2721 AppSerApLeuSerLeuGluGluAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
DB 6787 GACAGTGAACCTCTCTTGAAGAACAACACGAGATGGCTCTTATGCTTCAACCAACATCA 6846  
QY 2741 AppSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
DB 6847 GACAGTGAAG 6906  
QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGluGluGluGluGluGluGluGluGluGluGlu 2780  
DB 6907 TGGGATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6966  
QY 2781 GilyGlyProGlyProGlyValAspAlaProTrpProGlyAspPheGlyThrThrAlaGlyGlu 2800  
DB 6967 GGGGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7026  
QY 2801 SerSerGlyAsnGlyAlaProGluGluGluGluGluGluGluGluGluGluGluGluGlu 2820  
DB 7027 AGTAGTGCAACGGGGCCCTTAGAGAGCGGCTCGGAGAGATGAGATGCTGCTGCTGCA 7086  
QY 2821 GluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLeuGlyLeuLeuGly 2840  
DB 7087 GAGGGGCTCCCTAGGCGCCCTTCAGAGCTCTTCTGCGCCAGCTCCAAAGGATCCTTAG 7146  
QY 2841 LysLysGlyLeuProThrThrLeuSerGlyLeuSerSerLeuLeuGluGluGluGluGlu 2860  
DB 7147 AAGAGTGTCTGCGCCCAACATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7206  
QY 2861 CysThrGlySerSerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
DB 7207 TGCAAGGGTCTTCCCGGGGCTCTCCCTAGTAGAGGAGCGGGGGGGGGGGGGGGGGGGGG 7266  
QY 2881 ArgProProProArgGlnSerLeuGlnGluGluGluGluGluGluGluGluGluGluGlu 2900  
DB 7267 CGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7326  
QY 2901 SerLeuValGlyThrValAspGluAspSerSerGlySerGluPheLeuPheLeuAsn 2920  
DB 7327 AGCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7386  
QY 2921 PheLeuHis 2923  
DB 7387 TTCTCGCAT 7395

RESULT 8  
ABA19447  
ID ABA19447 standard; DNA; 2332 BP.  
XX ABA19447;  
AC  
XX 23-JAN-2002 (first entry)  
DE Human nervous system related polynucleotide SEQ ID NO 11778.  
XX  
KW Human; nocitropic; neuroprotective; cytostatic; dermatologic; vitruclide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerrary;  
KW antiparkinsonian; antistickling; antianemic; antiarthritic; cancer;  
KW antineumatic; hepatocytic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
XX  
XX 16-AUG-2001.  
XX

PF 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231122.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234974.  
PR 25-SEP-2000; 2000US-0234977.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242321.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,

PT useful for preventing, diagnosing and/or treating nervous system  
PR cancers and metastases -  
XX  
XX  
PS Disclosure; SEQ ID NO 11778; 1701bp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABAI1004-ABAI21534) and proteins  
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
SQ Sequence 2332 BP; 505 A; 669 C; 665 G; 493 T; 0 other;  
Alignment Scores:  
Pred. No.: 0 Length: 2332  
Score: 776.00 Matches: 776  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.55% Indels: 0  
DB: 22 Gaps: 0  
US-09-916-849A-3 (1-2923) x ABAI9447 (1-2332)  
QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
DB 3 ACCCAAGCTTCAGAGGTCACGCGCAGAGCAACGCGCAGATGCCCGCAGAGTCCCTGGCT  
QY 274 ThrLeuThrIleLeuValThrAspThrAspAspHisAspProValPheGluGlnGlnGlu 293  
DB 63 AACTCAACCATTTGTGTTAGTACACACATGACATGACATGACATGTTGAGCAGCAGAG 122  
QY 294 TyrLysGluSerLeuArgLysAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
DB 123 TACAAGAGAGAGCTCAGGAGAGAACTGAGAGTGTGCTATGAGTCTCACTCAGAGGCC 182  
QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluLysGly 333  
DB 183 ACGAGATGGATGATCCCTCCCATGCAATATTCGTACCGCTCTGAGAGGGGTCTGGG 242  
QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
DB 243 GGCAGCCCTCTGAAGCTTTGAGATCAACCTCCTTGAGGTATTCGAACCCGTGGC 302  
QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
DB 303 CCTGTGATCGGAGAGAGGTGAATCTTACGATCGAGCGTGAAGGCAAGTGAACAGGGT 362  
QY 374 ArgAspProGlyProArgSerThrThrAlaValPheLeuSerValGluAspAspAsn 393  
DB 363 CGGAGCCCGGTCTTCGAGATCAACACCGCTGTCTTCCTTCTGTGTGAGGATATCAAT 422  
QY 394 AspAsnAlaProGlnPheSerGluLysArgTyrValValGlnValArgGluAspValThr 413  
DB 423 GATATGCCCCCAATTATGATGAGAGCGCTATGTGTCTCAGGTAGAGGAGATGTGACT 482  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlyGlySerAsnAlaVal 433  
DB 483 CCAAGGAGCCCGAGTCTCGAGTCAACGCTCGATTCAGACAGAGGAGGAGCATCCGCG 542  
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGluGlnPheTyrLeuAspAlaGlnThr 453







CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells, AAK64703  
CC to AAK67694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
S0 Sequence 2332 BP, 505 A, 665 C, 665 G, 493 T, 0 other;

**Alignment Scores:**

Pred. No.:	0	Length:	2332
Score:	776.00	Matches:	776
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	26.55%	Indels:	0
DB:	22	Gaps:	0

US-09-916-849A-3 (1-2923) x AAK83060 (1-2332)

QY	254	ThrHisValPheArgValThrIaIaAspHisIleGluMetProArgArgSerAlaIleVal	273
Db	3	ACCACGGCTTCAGGGGTACGGGGGAGGACCAACGGGATCCCCGAGAAAGTCCCTGGCT	62
QY	274	ThrIleuThrIleLeuValThrAspThrAspAspHisAspProValPheGluGlnGlnIle	293
Db	63	ACACTCACCATCTTGTTAGTACTGACCAACATGACCAATGACCACTGTGTCTGTTCAGACGACGAG	122
QY	294	TyrTyrGluSerIleuArgGluAsnIleuGluValGlyTyrGluValIleuThrValArgAla	313
Db	123	TACACAGGAGAGCTCAGGAGGAGAACTGAGGGTGTGCTTAGAGGTCTCACTGCAAGGCC	182
QY	314	ThrAspGlyAspAlaProProAsnIleAsnIleLeuTyrArgIleuIleuGluSerGly	333
Db	183	ACGATGGTGAATGCCCTCCATGACCAATATCTGTACCGCTGTGTGAGAGGGGTGTGG	242
QY	334	GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly	353
Db	243	GCGAGCCCTCTGAAGTCTTTGAGATCGACCTCGCTCGGGGTGATCCGAACCCCTGGC	302
QY	354	ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly	373
Db	303	CCTGGATTCGGGAAGAGGTGAAATCTTACAGCTGACGGTGAAGGCAATGACCAAGGCT	362
QY	374	ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn	393
Db	363	CGGACCCGGGCTCTCGAGATGACCAACGCCGTGTTCTCTTCTGTGTGAGATGACAAAT	422
QY	394	AspAsnAlaProGlnPheSerGluTyrAspArgTyrValAlaGlnValArgGluAspValThr	413
Db	423	GATATGCCCCCACAATTATGTAAGAAGCCCTATGTGTCTCAAGTGAAGGAGATGTGACT	482
QY	414	ProGluAlaProValIleuArgValThrAlaSerAspArgAspLysSerAsnAlaVal	433
Db	483	CCAGGGGGCCAGTACTCCGAGTCCACAGCTCGGATCCGAGACAAAGGGAGCAATGCCGTG	542
QY	434	ValHisTyrSerIleMetSerGlyAsnAlaAspGlyGlnPheTyrIleuAspAlaGlnThr	453
Db	543	GTGACATATGATCATATGATGTGGCAATGTCGGGACAGTTTATCTCGATGCCCAAGACT	602
QY	454	GlyAlaIleuAspValValSerProIleuAspTyrGluThrThrTyrGluIuTyrThrLeuArg	473
Db	603	GGAAGCTCGAATGTGTGAGCCCTTTGACTATGAGACCAACAGAGATGACCTACGG	662
QY	474	ValArgAlaGlnAspGlyGlyArgProProIleuSerAspValSerGlyIleuValThrVal	493

Db	663	GTGGCGAGCAAGAGATGGTGGCCCTGCCCACTCTCTAATGTCTCGCTTGTCAGCTA	722
Qy	494	GlnValIleuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr	513
Db	723	CAGGCTCCGGATATCAACAGCAATCCCCCACTTCCTGTCAGCAACCCCTTCCAGGCTACT	782
Qy	514	ValIeuGluSerValProIeuGlyTyrtIeuValIeuHisValGlnAlaIleAspAlaAsp	533
Db	783	GTCCCTGGAGAGGCTCCCTTAGCTACTGGTTCTCCATGTCCAGGCTATTCAGACCTGAT	842
Qy	534	AlaGlyAspAsnAlaArgIeuGlnTyrtArgIeuAlaGlyValGlyHisAspPheProPhe	553
Db	843	GCTGTGTGCAATATGCCGCTGGAAATACCGGCTTGCTGGGGGTGGACATGACTTCCCTTC	902
Qy	554	ThrIleAsnAsnGlyThrGlyTyrtProIleSerValAlaAlaGlyIleuAspArgGlnGluVal	573
Db	903	ACCATCAACAAATGGCAAGGCTGATCTCTGTGGCTGTGAATGGACCGGAGGAAGTT	962
Qy	574	AspPheTyrtSerPheGlyValGlyAlaArgAspHisGlyTyrtProAlaIleuThrAlaSer	593
Db	963	GATTTCTACAGTTTGGGGTATGAAGCTGGAACCAATGGCACTTCAGCACTACTGCTCTCG	1022
Qy	594	AlaSerValSerValThrValIleuAspValAsnAspAsnAsnProThrPheThrGlnPro	613
Db	1023	GCCATGTGCAAGTGAACGTCTCTGATGTCAACGACAAATTCCAACCTTTAACCAACCA	1082
Qy	614	GluTyrtThrValArgIleuAsnGluAspAlaAlaValGlyTyrtSerValValThrAlaSer	633
Db	1083	CAGTTCACAGATGCGGCTCAATGAGATGCAAGCTGGGCAACAGGTGTGTGCGGTGCA	1144
Qy	634	AlaValAspArgAspAlaHisSerValIleThrTyrtGlnIleThrSergIlyAsnThrArg	653
Db	1143	GCTGTGGACCGTGATGCTCATAGTGTCACTACCAAGATCAACAGTGGCAATCTCGA	1202
Qy	654	AsnArgPheSeriIleThrSeriGlnSergIlyGlyGlyIleuValSerIleuAlaIleuProIeu	673
Db	1203	AACCGCTTCTCATCAACAGCCAAAGTGTGTGGCTGGATCCCTTGCCCTGCACTG	1262
Qy	674	AspTyrtIleuAsnGluArgGlnTyrtValIleuAlaValThrAlaSerAspGlyTyrtArgGln	693
Db	1263	GACTACAAACTTGAGCGGCACTAATGTGTGGCTGTACCGGCTCGATGGCACTCGGAG	1322
Qy	694	AspThrAlaGlnIleValIleAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	1323	GACACGGGCAAGATGTGTGTGAATGTCAACCCACCCCAACCCATCGTCTGTCTTTCAG	1382
Qy	714	SerSerHisTyrtThrTrpValAsnValAsnGluAspArgProAlaGlyTyrtThrValIleu	733
Db	1383	AGCTCCCACTTACAGTGAATGTTAATGAGAACCGGCGGCGAGGACCAACGATGGTGTG	1442
Qy	734	IleSerAlaThrAspGluAspThrGlyGlyIleuAsnAlaArgIleThrTyrtPheMetGluAsp	753
Db	1443	ATCAGCGGCACGGATGAGGACACAGGTGAGAAATCCCGCATCACTTATGAGAGGAC	1502
Qy	754	SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluIleu	773
Db	1503	AGCATCCCCCACTTCCGATGTGATGACAGCAACGGAGGCTGTGCACACCACAGGCTGAAGTG	1562
Qy	774	AspTyrtGluAspGlnValSerTyrtThrIleuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	1563	GACTACGAGAGCCAAAGTGTCTTACACCTTGGCCATTACTGTCGGGACAAATGGCAATCCC	1622
Qy	794	GlnIlySerAspThrThrTyrtIleuGlnIleIleuValIleuAspValIleuAspAsnAlaPro	813
Db	1623	CAGAAAGTCCGACACCACTACCTGGAGATCTGTGTGAACGACGTGAATGACAAAGCCCT	1682
Qy	814	GlnPheIleuAspAspSerTyrtGlnGlySerValTyrtGluAspValIleProPheThrSer	833
Db	1683	CAGTTCCTGGAGACTCTTACAGAGGAGAGTGTCTAATGAGATGTGCCAACCCTTCACTAAGC	1742
Qy	834	ValIeuGlnIleSerAlaThrAspArgAspSergIlyIleuAsnGlyArgValPheTyrtThr	853
Db	1743	GTCCCTGAGATCTCAGCACTGATCGATTCGATTCGACTTATGCAAGGATCTTCTTACACC	1802

QY 854 PhcGInGlyGlyAspAspGlyAspPheIleValIuSerThsSergIylIleVal 873  
DB 1803 TTCcAGAGAGcCGAcGATGAGAcCGGTGACTTATTGTCACAGTCAGCATCTG 1862  
QY 874 ArgThrIeuArgArgIeuAspArgGluAsnValAlaGlnIyrValIeuAgaIaIyrAla 893  
DB 1863 CGAAGCGTACGAGAGCGGTGATTCAGAGAACGTGGCCAGATATCTTGGCGGATATGCA 1922  
QY 894 ValAspIleGlyMetProProAlaArgThrProMetGluValThrValThrValIeuAsp 913  
DB 1923 GTGGACAGAGGAGTGGCCCGACGCCGACACTATGGAATGACAGCTGCTTGAT 1982  
QY 914 ValAsnAspAsnProProValPheGluIleAspGluPheAspValPheValGluIleAsn 933  
DB 1983 GTCAATGACATATCCCTCTGTCTTGTAGCAGATGATGATGATGTTGTGGAAGAGAAC 2042  
QY 934 SerProIleGlyIleuAlaValAlaArgValThrAlaThrAspProAspGluIyrThrAsn 953  
DB 2043 AGCCCAATTGGGCTAGCGCGGTGTCACAGCCACTGACCCCGATGAGGACCACTCAT 2102  
QY 954 AlaGlnIleMetIyrGlnIleValIleGluIleAsnIleProGluValPheGlnIleAspIle 973  
DB 2103 GCCCAGATTATATGACAGATTTGTGAGGGGACATCCCTAGGCTCTTCCAGCTGGACATC 2162  
QY 974 PheSerGlyGluIeuThrAlaIeuValAspIleAspIyrGluAspArgProGluIyrVal 993  
DB 2163 TTCTCCGGGAGCTGACAGCCCTGTGTAGACTTACTGACGAGGACGGCCTGAGTACGTC 2222  
QY 994 IeuValIleGlnIleThrSerAlaProIeuValSerArgIleThrValHisValGluIle 1013  
DB 2223 CTGGTATCCAGGCGCAGTACGCTCTGTGTAGACCGGCTACAGTCCAGTCCGCTTC 2282  
QY 1014 IeuAspArgAsnAspAsnProProValIeuGlyAsnPheGluIleu 1029  
DB 2283 CTGGACCGGATGACCAACCAACGAGTGTGGCACTTGAGATCCTT 2330

RESULT 10  
AAS31490  
ID AAS31490 standard; DNA; 2332 BP.  
XX  
AC AAS31490;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human DNA for a novel extracellular matrix protein, Seq ID No 569.  
XX  
KW Human; secreted extracellular matrix protein; ds; immunomodulatory;  
XX Anti-HIV; antineoplastic; antirheumatic; antisclerotic; cardiant; vascular;  
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;  
XX anticancer; immune/autoimmune disease; HIV infection; anaemia;  
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;  
XX cancers; hyperproliferative disorder; breast neoplasm; melanoma;  
KW Sezary syndrome; Gaucher's disease; neurological diseases;  
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;  
KW cardiac arrest; tachycardia; angina; infection; corneal infections;  
KW wound healing; immunogen; gene therapy; antisense; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO200155368-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01348.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0215647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226886.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI; 2001-465572/50.  
 XX  
 PT Nucleic acid molecules encoding human secreted extracellular matrix  
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.  
 PT Alzheimer's and Parkinson's diseases and cancers -  
 XX  
 PS Claim 1; SEQ ID No 569; 577bp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules encoding  
 CC novel human secreted extracellular matrix proteins (SPs). The  
 CC polynucleotides and proteins are used to prevent, treat a medical  
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,

CC chickens or sheep. For example, disorders associated with decreased  
 CC expression of SPs. The SP polynucleotide or a vector expressing them may  
 CC be administered to treat diseases by gene therapy. Antisense molecules  
 CC may be administered to down regulate expression of SPs by binding with  
 CC the cells own genes and preventing their expression. The polynucleotides  
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be  
 CC used as antigens to produce antibodies and to identify modulators  
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and  
 CC antagonists may also be used to down regulate expression and activity of  
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.  
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV  
 CC (human immunodeficiency virus) infections, anemia, rheumatoid arthritis  
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.  
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and  
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac  
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses  
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses  
 CC include wound healing, maintenance of organs before transplantation,  
 CC support of cell culture of primary tissues, modulation of for example

## Alignment Scores:

Pred. No.: 0 Length: 2332  
 Score: 776.00 Matches: 776  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.55% Indels: 0  
 DB: 22 Gaps: 0

US-09-916-849a-3 (1-2923) x AAS31490 (1-2332)

QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
 DB 3 ACCACGCTTCAAGGTCACGGCCGACAGACCAACGCGATGCCCGACCAAGTCCCTGGCT  
 QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlu 293  
 DB 63 ACCTCACCATCTTGTTACTGACACCAATACATACACCTGTGTTGACACACAGAG  
 QY 294 TyrIysGlnSerLeuAsnGluAsnLeuGlnValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAGAGAGACCTCAAGAGAGAACCTGAGGTTGCTATAGAGTCTACGTCAGGGCC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlySerGly 333  
 DB 183 ACGATGTCATGTCCTCCCAATGCCAATATTCGTACCCCTGCTGAGGGGTCTGGG 242  
 QY 334 GlySerProSerGluValPheGlnIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GGCAGCCCTCTGAAGTCCTTGAATGATGACCTGCTGAGGATCGAACCCTGGGC 302  
 QY 354 ProValAspArgGlnGluValGlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGly 373  
 DB 303 CCTGTGATCGGAGAGAGGTGATCTTCAAGTGAAGGTGAAGGCAAGTGAACCAAGGT 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393  
 DB 363 CGGAGCCCGGGTCTCCGAGTACCAACAGCCGCTGTTTCTTCTTCGTGAGAGATGACAT 422  
 QY 394 AspAsnAlaProGlnPheSerGlnValArgTyrValValGlnValArgGluAspValThr 413  
 DB 423 GATATAGCCCTCCGATTAGTGAAGAGACGCTATGTGCTCAGTGAAGAGAGATGACT 482  
 QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
 DB 483 CCAAGGGGCCAGTACTCCAGTCAACGCTCGATGAGACAGAGGAGCAAGCCGCTG 542  
 QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
 DB 543 GTGCACTATAGCAATCAGATGAGGCAATGCTCGGAGCAAGTTTATCTGAGATCCAGACT 602  
 QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGlnThrThrIlysgIuTyrThrLeuArg 473

Db 603 GGAAGCTCTGGATGTGTGAGCCCTCTTGACATAGACGACCAAGAGTACACCTTCAGG 662  
 Qy 474 ValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
 Db 663 GTCCGAGACAGAGATGGTGGCCCTCCCACTCTTAATGTCTCTGGCTTGGAGACGTA 722  
 Qy 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
 Db 723 CAGGCTCTGGATATCAACGACCAATGGCCCCCATCTGCTGACGACACCCCTTCCAGGCTACT 782  
 Qy 514 ValLeuGlnSerValProLeuGlyValLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
 Db 783 GTCTGGAGAGGCGTCCCTTAGGCTACCTGTCTCTCATGTCCAGGCTATCGACGCTGAT 842  
 Qy 534 AlaGlyAspAsnAlaArgLeuGlyValArgLeuAlaGlyValGlyHisAspPheProPhe 553  
 Db 843 GCTGGAGACATGCCCGCCCTGGAAATACCGCTTGTCTGGGGTGGGACATGACTTCCCTTC 902  
 Qy 554 ThrIleAsnAsnGlyThrGlyValProIleSerValAlaAlaGlyLeuAspArgGlyVal 573  
 Db 903 ACCATCAACATGGCACAGGCTGGATCTCTGTGGCTGTACCTGGACCTGGAGGGAAGTT 962  
 Qy 574 AspPheTyrSerPheGlyValGlyAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
 Db 963 GATTTCTACAGCTTTGGGGTAGAAGCTCGAGACCATGGCACTCGACGACTCACTGCTCG 1022  
 Qy 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613  
 Db 1023 GCCAGGTGACCGGTGACTGTCTGGATGTCAACGACCAATCAACTTATCCACACCA 1082  
 Qy 614 GlnTyrThrValArgLeuAsnGlyAspAlaAlaValGlyThrSerValValThrValSer 633  
 Db 1083 GAGTACACAGATGGCGCTCATGTAGATGAGATGACGCTGGGCAACGCGGTGGAGCGGTGCTCA 1142  
 Qy 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653  
 Db 1143 GCTGTGACCGGTATGCTCATAGTGTATCATCACTTACCAATTCACAGTGGCAATCACTCA 1202  
 Qy 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673  
 Db 1203 AACCGCTTCCATTCACCAAGGTGGTGGGCTGGTATCCCTTGGCCCTGACACG 1262  
 Qy 674 AspTyrIleLeuGlnIleThrGlnIleThrValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
 Db 1263 GACTTACAAACTTGAAGCGGAGTATGTGTGCTGTTACCCGCTCCATGCGCACTCGGCA 1322  
 Qy 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713  
 Db 1323 GACACGGCACAGATGTGTGGAATGTCAACGACGCAACCCATGCTGCTTTCAG 1382  
 Qy 714 SerSerHisTyrThrValAsnValAsnGlyAspArgProAlaGlyThrThrValValLeu 733  
 Db 1383 AGCTCCCACTATATCAAGTAATGTTAATGAGAGACCGGCGGACGACCAACGAGTGGCTG 1442  
 Qy 734 IleSerAlaThrAspGlyAspThrGlyLeuAsnAlaArgIleThrTyrPheMetGlyAsp 753  
 Db 1443 ATCAGCGCCACGAGTGAAGCACAGGAGTAAGCCGCAATCACTTCAATGAGGAGAC 1502  
 Qy 754 SerIleProGlnPheArgIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGlyLeu 773  
 Db 1503 AGCATCCCGCCAGTTCGCCATCGATGACGACACCGGGGCTCTCACACCCAGCTGAGCTG 1562  
 Qy 774 AspTyrGlnAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
 Db 1563 GACTACCAACACCAAGTGTCTTACACCTGCGCAATTACTCTCGGAGCAATGACATTTCCC 1622  
 Qy 794 GlnIleSerAspThrThrTyrIleGlnIleLeuValAsnAspValAsnAspAsnAlaPro 813  
 Db 1623 CAGAGTCCACACACCACTACCTGAGATCTGGTGAACACGTTGAATGACATGGCCCT 1682  
 Qy 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGlnAspValProProPheThrSer 833  
 Db 1683 CAGTCTTCGAGACCTCTCAACAGGGGAGTGTGTATGAGATGTGGACACCTTCACTAGC 1742

Qy 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853  
 Db 1743 GCTCTGACAGATTCAGCCACTGATGTGATCTTGACCTTAATGGACGGCTTCTTACACC 1802  
 Qy 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGlnSerThrSerGlyIleVal 873  
 Db 1803 TTCCAAAGAGGGGAGAGATGAGACCGGTATCTTATGTGTAGTCCACGTCAGGCAATCGTG 1862  
 Qy 874 ArgThrLeuArgArgLeuAspArgGlyAsnValAlaGlnIleThrValLeuArgAlaTyrAla 893  
 Db 1863 CGAAGCTTACGAGGCTGGATGTGAGAGAACGTGGCCAGTATGTCTTGGCGGCAATGGA 1922  
 Qy 894 ValAspIleGlyMetProProAlaArgThrProMetGlyValThrValThrValLeuAsp 913  
 Db 1923 GTGGACAGGGGATGCCCGCCAGCCGACACTTAAGAAAGTACAGTCACTGTGTGGAT 1982  
 Qy 914 ValAsnAspAsnProProValPheGlnGlnAspGlyPheAspValPheValGlnGlnAsn 933  
 Db 1983 GTGAATGACAAATCCCGCTGTGTGAGCAGATGATGTGTGTGTGGAAGAGAAC 2042  
 Qy 934 SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlnGlyThrAsn 953  
 Db 2043 ACCCCCATTTGGCTTACCGGTGGCCCGGTCAAGCCCACTGACCCCGATGAAAGGCAACAT 2102  
 Qy 954 AlaGlnIleMetTyrGlnIleValGlnGlyAsnIleProGlnValPheGlnLeuAspIle 973  
 Db 2103 GCCAGATTATATGATACAGATTGTGAGAGGCAACATCCCTGAGGTCTTCCAGCTGACATC 2162  
 Qy 974 PheSerGlyGlnLeuThrAlaLeuValAspLeuAspTyrGlyAspArgProGlnTyrVal 993  
 Db 2163 TTCTCTCGGGAGACTTACAGCCCTGTGTAGACTTACAGAGAGACCGGCTGAGTACGTC 2222  
 Qy 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1013  
 Db 2223 CTGTGTATCCAGGCCACAGTCACTCTCTGTGTGAGCCGGGCTTACAGTCCAGCTCCGCTC 2282  
 Qy 1014 LeuAspArgAsnAspAsnProProValLeuGlyAsnPheGlnIleLeu 1029  
 Db 2283 CTTGACCCGCAATGACCAACCCACAGTGGGCACTTGTGAGATCTT 2330  
 RESULT 11  
 ABO66814  
 ID ABO66814 standard; DNA; 2332 BP.  
 AC  
 XX  
 XX  
 DT 23-AUG-2002 (first entry)  
 XX  
 XX  
 DE Human polynucleotide SEQ ID NO 569.  
 XX  
 XX  
 KW Human; noctropic; neuroprotective; cytoskeletal; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antischizoid; antianemic; antidiabetic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 OS Homo sapiens.  
 XX  
 XX  
 PN US2002042386-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 XX  
 PF 17-JAN-2001; 2001US-0764870.  
 XX  
 PR 31-JAN-2000; 2000US-179065P.  
 PR 04-FEB-2000; 2000US-180628P.  
 PR 28-JUN-2000; 2000US-214886P.  
 PR 07-JUL-2000; 2000US-216647P.  
 PR 07-JUL-2000; 2000US-216880P.  
 PR 11-JUL-2000; 2000US-217487P.

PR 11-JUL-2000; 2000US-217496P.  
 PR 14-JUL-2000; 2000US-218290P.  
 PR 26-JUL-2000; 2000US-22063P.  
 PR 26-JUL-2000; 2000US-22064P.  
 PR 14-AUG-2000; 2000US-224518P.  
 PR 14-AUG-2000; 2000US-224519P.  
 PR 14-AUG-2000; 2000US-225267P.  
 PR 14-AUG-2000; 2000US-225268P.  
 PR 14-AUG-2000; 2000US-225270P.  
 PR 14-AUG-2000; 2000US-225477P.  
 PR 14-AUG-2000; 2000US-225757P.  
 PR 14-AUG-2000; 2000US-225758P.  
 PR 22-AUG-2000; 2000US-226688P.  
 PR 30-AUG-2000; 2000US-228924P.  
 PR 01-SEP-2000; 2000US-229287P.  
 PR 01-SEP-2000; 2000US-229343P.  
 PR 01-SEP-2000; 2000US-229344P.  
 PR 01-SEP-2000; 2000US-229345P.  
 PR 05-SEP-2000; 2000US-229509P.  
 PR 05-SEP-2000; 2000US-229513P.  
 PR 08-SEP-2000; 2000US-231413P.  
 PR 21-SEP-2000; 2000US-234223P.  
 PR 21-SEP-2000; 2000US-234274P.  
 PR 25-SEP-2000; 2000US-234997P.  
 PR 27-SEP-2000; 2000US-235834P.  
 PR 29-SEP-2000; 2000US-236327P.  
 PR 29-SEP-2000; 2000US-236367P.  
 PR 29-SEP-2000; 2000US-236368P.  
 PR 29-SEP-2000; 2000US-236369P.  
 PR 29-SEP-2000; 2000US-236370P.  
 PR 02-OCT-2000; 2000US-236802P.  
 PR 02-OCT-2000; 2000US-237037P.  
 PR 02-OCT-2000; 2000US-237038P.  
 PR 02-OCT-2000; 2000US-237039P.  
 PR 02-OCT-2000; 2000US-237040P.  
 PR 13-OCT-2000; 2000US-237045P.  
 PR 20-OCT-2000; 2000US-239335P.  
 PR 20-OCT-2000; 2000US-240960P.  
 PR 20-OCT-2000; 2000US-241785P.  
 PR 20-OCT-2000; 2000US-241809P.  
 PR 01-NOV-2000; 2000US-244617P.  
 PR 17-NOV-2000; 2000US-249299P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (RUBB/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 DR WPI: 2002-470713/50.  
 XX  
 PT New nucleic acid encoding human proteins, useful for diagnosis,  
 PT treatment and prevention of e.g. osteoporosis, also related  
 PT polypeptides and antibodies -  
 XX  
 ES Disclosure; SEQ ID NO 569; 235bp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (AB06521-AB066785) and proteins  
 CC (ABP4746-ABP48110) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections. The present sequence is that of a  
 CC polynucleotide of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from USPTO at [seqdata.uspto.gov/sequence.html?docID=999909764870](http://seqdata.uspto.gov/sequence.html?docID=999909764870).  
 XX  
 SQ Sequence 2332 BP, 505 A; 669 C; 665 G; 493 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2332  
 Score: 776.00 Matches: 776  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.55% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-916-849a-3 (1-2923) x AB066814 (1-2332)  
 QY 254 ThrHsValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
 DB 3 ACCACGCTTCAGAGTCACGCGCAGAGCACGCGCATGCCCGACGAGTGCCTGCT 62  
 QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlu 293  
 DB 63 ACACCTCACCATCTTGTTACTGACACCATATACCATACCTGTGTTCAGACAGCAGG 122  
 QY 294 TyrIysGluSerIleuArgIleuAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAGAGAGAGCTCAGAGGAGAACCTGAGGTTGCTATGAGTGTCTCATCTCAGGCGC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGly 333  
 DB 183 ACGATGCTGATGTCCTCCCAATGCCAATATTCGATCCGCTGCGAGAGGCTCTGGG 242  
 QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GCGAGCCCTCTGAAGCTTTGATGATGACCTCTGCTGGGAGTATCCAGCCCGTGGC 302  
 QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
 DB 303 CCTGTGATCGGAGAGAGTGAATCCTACACGATGAGCGTGAAGGACCAAGGAT 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaValPheLeuSerValGluAspAsn 393  
 DB 363 CGGAGCCCGGCTCTGAGAGTACCAAGCCGCTGTTTCTTCTTGAGAGATGACAT 422  
 QY 394 AspAsnAlaProGlnPheSerGluLysArgTyrValValGlnValArgGluAspValThr 413  
 DB 423 GATATATGCCCCCGATTTATGAGAGACCGCTATGTGTCAGAGTGAAGATGTACT 482  
 QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
 DB 483 CCAGGGGCCCCAGATCTCCGATGACAGCTCGGATGAGCAAGGAGGAGATCGGTG 542  
 QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
 DB 543 GTGCACATAGACATCAAGAGTGGCAATGCTCGGGGACAGTTTATCTGAGTCCCAACT 602  
 QY 454 GlnValLeuAspValValSerProLeuAspTyrGlnThrThrIlysgGluTyrThrLeuArg 473  
 DB 603 GGACCTCTGATGTGTGAGACCCCTTGTATGACAGACGACGAGATCACTTACGCG 662  
 QY 474 ValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
 DB 663 GTGGAGACACAGAGTGTGCGCCGCCCACTCTATATGTCTGTGCTGTGTACAGTA 722  
 QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
 DB 723 CAGGTCTTGATATCAACGACATGCCCCCATCTTGTGAGACCCCTTTCAGGCTACT 782  
 QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
 DB 783 GTCTTGAGAGAGCGTCCCTTAGGTACTGTGTTCTCATATGTCAGGCTATCGAGCTGAT 842

QY	534	AlaGlyAspAsnAlaArgLeuGlnTyrArgLeuAlaGlyValGlnIleAspPheProPhe	555
Db	843	GCTGTGACAAATGCGCGCTGGATACCGCTTGTCTGGGGTGGGACATGACTTCCCTTC	902
QY	554	ThrIleAsnAsnGlnTyrGlnTyrIleSerValAlaAlaGluLeuAspArgGluGluVal	573
Db	903	ACCATCAACAAATGGCACACAGCTGGATCTCTGTGCTGCTGAACCTGGACCCGGAGAAAGTT	962
QY	574	AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer	593
Db	963	GATTCTTACAGCTTTGGGTGAAAGCTCGAAGCAATGGCACTCCAGCACTACGTCCCTCG	1022
QY	594	AlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnPro	613
Db	1023	GCCAGTGTCAACGTGACTGTCTCTGGATGTCAACGCAAAATTCACCACTTTAACCAACA	1082
QY	614	GlnTyrThrValAlaGluLeuAsnGluAspAlaAlaValGlnTyrSerValValThrValSer	633
Db	1083	GAGTACACAGTGGCGCTCAATGAGAGATGCACCTGTGGGCACACAGCTGTATACAGTGTCA	1142
QY	634	AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg	653
Db	1143	GCTGTGACCGCTGATGCTCAATAGTGTATCACTACCAAGATCAACAGTGGCAATCTCGA	1202
QY	654	AsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	1203	AAACGGCTTCTCCATCACACCAAGCAAAAGTGTGTGGTGGCTGTATCCCTTGCCCTCCACTG	1262
QY	674	AspTyrIleValLeuGlnTyrValLeuAlaValThrAlaSerAspGlyTyrThrArgGln	693
Db	1263	GACTACAACTTGAGCGGACAGTATGTGTGCTGTATCCGCTCCGATGGGACCTCGGAG	1322
QY	694	AspThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	1323	GACACGGCACAGATTGTGTGAATGTACCCAGCGCAACCACTTCGTCTGTCTTTCAG	1382
QY	714	SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyTyrThrValValLeu	733
Db	1383	AGCTCCCACTATACAGTAATGTTAATAGACACGGCGGACGACCAACGCTGGTGTG	1442
QY	734	IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp	753
Db	1443	ATCAGCGCCACGAGTACAGACACAGGTAGAAATGCCGCACTACCTTACATGAGGAC	1502
QY	754	SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu	773
Db	1503	AGCATCCCGCCAGTTCGCGATCGATGCATGACACACGGGGGCTGTACCAACCGCTGAGCTG	1562
QY	774	AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	1563	GACTACGAAGACCAAGATGTCTTACACCCCTGGCACTTACGTCTCGGGAACAAATGGCAATCCC	1622
QY	794	GlnTyrSerAspThrThrTyrIleGlnIleLeuValAsnAspValAsnAspAsnAlaPro	813
Db	1623	CAGAAAGTCGACACCACTTACTGTGAATCTCGTGGAAACAGCTGAATGACATGCCCT	1682
QY	814	GlnPheLeuAspAspSerTyrGlnTyrSerValTyrGluAspValProProPheThrSer	833
Db	1683	CAGTTCCTCGCAGACTCTCTTACACAGGCGAGTGTCTATGAGATGTGCCACCTTCACTAGC	1742
QY	834	ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr	853
Db	1743	GTCTTCGACGATTCACGCCATGATCTGATTCGTGAATGGAAGGAGCTTCTTCAACC	1802
QY	854	PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal	873
Db	1803	TTCCAAAGAGCGCAGATGAGAGACGCTGACTTTATTTGAGTCCACGTCAAGCAATCGTG	1862
QY	874	ArgThrLeuAspArgLeuAspArgGluLeuValAlaGlnTyrValLeuArgAlaTyrVal	893
Db	1863	CGAAGCTTACGAGAGCTGTGATTCGAGAGAACTGTGCCAGTATGCTTTCGGGGCAATATGCA	1922

QY		894	ValAspLyseGIymetProPcoAlAargThrPrometGluValThrValThrValLeuAsp	913
Dd		1923	GTCGACAAAGGGGATGCCCCCAAGCCCGCACACTATGGAAGAATGCACATCTGTTCGAT	1982
QY		914	ValAsnAspAsnProProValPhegluglnAspgIuPheAspValPheValGluglunsn	933
Dd		1983	GTAATAACAATCCCCCTGCTTCTTGAGCAGAGAAGATTGATGTGTGTGGAGAGAAC	2042
QY		934	SerProtllegIleuAlaValAlaArgValThrAlaThrAspProAspGluGIYThrAsn	953
Dd		2043	AGCCCCATTGGGCTTAGCCGTTGGCCCGGATCACACCCTAACCCTCCCATGAAAGCAACAT	2102
QY		954	AlaGlnIleMetYrrgInIlleValgluglyAsnIleProGluValPheGlnLeuaspIle	973
Dd		2103	GCCCAAGATTATGTACCAAGATTGTGAGAGGCACAATCCCTGAGGCTTCCAGCTGACATC	2162
QY		974	PheSerclVgluIeuThrAlaLeuValAspLeuAspYrrgluAsparqProGluTYrVal	993
Dd		2163	TTCCTCCGGGAGCTGACAGCCCTGGTAGACTTGACTGACACAGAGACCGGCTGAGTACGTC	2222
QY		994	IeuValIIegInAlaThrSerAlaProLeuValSerArgAlaThrValHlsvalArgIeu	1013
Dd		2223	CTGGTATTCAGAGCGCAAGTCAAGCTCCCTGTGTAGCGGGGTACAGTCCAGCTCGGCTC	2282
QY		1014	LeuAspArgAsnAspAsnProProValIleuglyAsnPhegluIleu	1029
Dd		2283	CTTGACCGCAATGACAAACCCACAGTGTGGGCAACTTGTAGATCCTT	2330
<b>RESULT 12</b>				
ABT31943	ID	ABT31943	standard; DNA; 2391 BP.	
XX	AC	ABT31943;		
XX	DT	01-MAY-2003	(first entry)	
XX	DE		Human breast cancer / ovarian cancer related coding sequence #50.	
XX	KW		Human; gene; ds; cytoblastic; breast cancer; ovarian cancer.	
XX	OS		Homo sapiens.	
XX	PN	WO2003000012-A2.		
PD	XX	03-JAN-2003.		
PE	XX	21-JUN-2002; 2002MO-US19773.		
PR	XX	21-JUN-2001; 2001US-300159P.		
PR	XX	27-JUN-2001; 2001US-301351P.		
FA	XX	(MTLL-) MILLENNIUM PHARM INC.		
XX	XX	Velby OP;		
XX	DR	WPI; 2003-267848/26.		
XX	DR	P-PsDB; ABJ37074.		
PT	XX		Determining the presence of breast cancer in an individual, involves	
PT	XX		using specific polynucleotide markers -	
PS	XX		Disclosure; Page 218; 233pp; English.	
CC	XX		The invention comprises a method for assessing whether a patient is	
CC	XX		afflicted with breast cancer or ovarian cancer. The method involves the	
CC	XX		use of specific DNA markers. The method of the invention is useful in the	
CC	XX		detection and treatment of ovarian and breast cancer. DNA sequences	
CC	XX		ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins.	
SO		Sequence 2391 BP; 460 A; 810 C; 679 G; 442 T; 0 other;		
<b>Alignment Scores:</b>				
Pred. No.:		0	Length:	2391



Score:	729.00	Matches:	729
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatch:	0
Query Match:	24.94%	Indels:	0
DB:	25	Gaps:	0

[illegible]

Db	979	GGAAAGCCCTGACTCTGTGCTGCTCTCCACTATATACAGCTCATCTGGAGTTTGTGCTGC	1038
Qy	2526	ProValAlaPhealAvalSerMetSerValPheLeuTyrlleuAlaAlaArgAlaSer	2545
Db	1039	CCGGAGGCTTTTGCCTCTCTGATGATGATGCTTCTCTGTAACATCTGGGGGCGGCGCTCC	1098
Qy	2546	CysAlaAlaGlnAargGlnGlyPheGlnLulysGlyProValSerGlyLeuGlnProSer	2565
Db	1099	TGTGTGTGCCCAAGGCGGAGGCTTTGAAGAAAGAAAGTCTGTCTCGGGCGTGGACCTTCC	1158
Qy	2566	PheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSer	2585
Db	1159	TTTCGCCGCTCTCTGCTGCTGAGGCCACCGTGGCTGGGCACTGCTCTCTGTACAGC	1218
Qy	2586	AspThrLeuLeuPheMetIleTyrlleuPheAlaThrCysAsnCysIleGlnGlyProPheIle	2605
Db	1219	GACACCCCTCTCTTCCACTACCTCTTTGTCTACCTGGCAATTGCAATCCAGGGCGCTTATC	1278
Qy	2606	PheLeuSerTyrlValValLeuSerGlyValAlaArgValAlaLeuLysLeuAlaCysSer	2625
Db	1279	TTCTCTCTCTAATGATGATGCTTAAAGAAAGGCTCGGAAGAGCACTCAAGCTTGGCTGAGC	1338
Qy	2626	ArgLysProSerProAspProAlaLeuLeuThrTrpLysSerThrLeuThrSerSerTyrlAsn	2645
Db	1339	CGCAAGCCCAAGCCCTGACCTGCTCTGACCAACCAAGTCCACCCGACTGCTGCTCTTAAC	1398
Qy	2646	CysProSerProTyrlAlaAspGlyArgLeuTyrlGlnProTyrlGlyAspSerAlaGlySer	2665
Db	1399	TGCCCCAGCCCTTAAGAGATGGGGCGGCTGTACAGGCCCTACAGGACTGGCGCGCTCT	1458
Qy	2666	LeuHisSerThrSerAspSerGlyLysSerGlnProSerTyrlIleProPheLeuLeuArg	2685
Db	1459	CTGCACAGCAACCAATCCCTCGGGGCAAGTCAAGCCAGCTCAATCCCTCTTCTGCTAGG	1518
Qy	2686	GlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeu	2705
Db	1519	GAGAGTCCCGACATGAAACCTTGGCCAAAGGGCGCCCTGTGGCGGGGAGATCCAGCGAGCTG	1578
Qy	2706	PheLeuGlnGlyGlnAAspGlnGlnHisAspProAspThrAspSerAspLeuSer	2725
Db	1579	TTCTCTGAAAGTCAAGACCAAGCAGCATGATCTCTGACAGCATCTCCGACATGACTCTTCC	1638
Qy	2726	LeuGlnAspAspGlnSerGlySerTyrlAlaSerThrHisSerSerAspSerGlnGln	2745
Db	1639	TTAAGAAACGACCAAGATGGCTCTCTAAGCTCTAACCACTCATCAGACAGTGGAGAGAA	1698
Qy	2746	GlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTrpAspSerLeuLeu	2765
Db	1699	GAAAGAGAGAGAGAAAGAGAGGCGGCTTCCCTGGAGAGAGAGGCTGGAGATGACTGCTG	1758
Qy	2766	GlyProGlyAlaGlnAAspLeuProLeuHisSerThrProLysAspGlyGlyProGlyPro	2785
Db	1759	GGGCTGGAGAGAGAGACTGGCCCTCTGACAGTACTCCAGAGATGGGGGCCCAAGGCGCT	1818
Qy	2786	GlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlnSerSerGlyAsnGly	2805
Db	1819	GGCAGAGCGCCCTGCGCAGAGACTTTGGAGACCAACAGCAAAAGAGTACTGTGCACAGGG	1878
Qy	2806	AlaProGlnGlnAAspGlyAlaAsnGlyAspAlaLeuSerArgGlnGlySerLeuGly	2825
Db	1879	GGCCCTGAGAGAGCGGCTGGCGGAGAAATGAGATGCCCTGTCTCGAGAGGGGCTCTTAGGC	1938
Qy	2826	ProLeuProGlySerSerAlaGlnProHisLysGlyTlleLeuLysLysCysLeuPro	2845
Db	1939	CCCCCTTCAGGCTCTTGTGCCAGGCTCAACAAAGGCACTCTTAAGAAAGATGTCTGCGCC	1998
Qy	2846	ThrIleSerGlyLysSerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSer	2865
Db	1999	ACCATCAGCGAAGAGACAGCTTCTGCGGCTCCCTTGGAGCAATGCAAGGGCTTCTCC	2058
Qy	2866	ArgGlySerSerAlaSerGlnGlnGlySerArgGlyGlyProProProArgProProArg	2885

Db 2059 CGGGGCTCCTCCGCTAGTGAGGCGAGCGGGGCGGCCCCCTCCGCCGCCACCGCCCGG 2118  
Qy 2886 GlnSerLeuGlnGlnGlnLeuAenGlyValMetProIleAlaMetSerIleGlyAlaGly 2905  
Db 2119 CAGAGCTCCAGAGAGCGAGCTGAGCGGGGTCAATGCCCATTCGACATCAAGGCGAGGC 2178  
Qy 2906 ThrValAspGluAspSerSerGlySer 2914  
Db 2179 ACGGTGATGAGAGACTGTCAGGCTCC 2205  
RESULT 13  
ABAI9448  
ID ABAI9448 standard; DNA; 2077 BP.  
XX ABAI9448;  
XX  
XX  
XX 23-JUN-2002 (first entry)  
DE Human nervous system related polynucleotide SEQ ID NO 11779.  
XX  
XX Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischizling; antianemic; antiarthritic; cancer;  
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
OS Homo sapiens.  
XX  
XX W0200159063-A2.  
XX  
XX 16-AUG-2001.  
PD  
PF 17-JAN-2001; 2001MO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198173.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216860.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231415.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239393.  
PR 13-OCT-2000; 2000US-0239397.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.



QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrlinleTherSerGlyAsnThrArg 653  
DB 1143 GCTGTGACCGTATGTCTCATATGTCATCATCCATCAAGTATCAACGTGCAATATCTCA 1202  
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673  
DB 1203 AACCGCTTCTCCATCAACAGCAAGAGTGTGGGTGTATCCCTTGTCCCTCCACTG 1262  
QY 674 AspTyrLeuGlnLeuArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
DB 1263 GACTCAAAATTTAGACGGCAGTATGTGTGGCTGTTCACCTCCATGCACTCCGGCAG 1322  
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrAlaArgProValIleGln 713  
DB 1323 GACACGGCAGACAGATTGTGTGAATGTCAACGACCCACACCATGTCTGTCTTTCA 1382  
QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu 733  
DB 1383 AGCTCCCATATACAGTAATGTTAATGAGGACCGCGGCGAGCAGCAGGAGTGTCTG 1442  
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753  
DB 1443 ATCAGCGCCACGATGAGACACAGGTGAGAAATCCCGCATCACTTCAATGAGAGAC 1502  
QY 754 SerIleProGlnPheArgIleAspAlaAspThrGlyValAlaValThrGlnAlaGluLeu 773  
DB 1503 AGCATCCCGCATGTCGATGATGACAGACGCGGGGTGTCAACCCAGGCTGAGCTG 1562  
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
DB 1563 GACTATGAACACAAAGTGTCTTACACCTGCGCATTACTCTCGGAGCAATGCGATTTCC 1622  
QY 794 GlnIleSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro 813  
DB 1623 CAGAAAGTCCAGACACCACTACCTGAGATCTGGTGAACAAGTGAATGCAATGCCCT 1682  
QY 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833  
DB 1683 CAGTTCCTCGAGAGTCTCTACAGGGCAGTGTATAGATGATGCGACCTTCATAGC 1742  
QY 834 ValIleGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValIlePheTyrThr 853  
DB 1743 GTCTTGAGATCTCAGCACTGATCGATTCGATTCGACTTAATGCGAGGTCCTTCTAC 1802  
QY 854 PheGlnIleGlyAspAspGlyAspGlyAspPheIleValGlnSerThrSerGlyIleVal 873  
DB 1803 TTCCAAAGAGCGCAGATGAGAGACGATGACCTTATTGTTCAGTCAAGCATGCTG 1862  
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuAlaIleTyrAla 893  
DB 1863 CGAAGCTTACGAGGCTGATCGAGAGAACGTGGCCCAATATGCTTGGCGGCAATGCA 1922  
QY 894 ValAspArgGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913  
DB 1923 GTGGACAGAGGGATGCCCCCAGCCGCAACCTATGAAATGACAGTCACTGTGTGGAT 1982  
QY 914 ValAsnAspAsnProProValIleGlnGlnAspGluPheAspValPheValIleGluAsn 933  
DB 1983 GTGAATGACATCCCTGCTTGTGACAGAGATGATGATGATGATGATGATGATGAT 2042  
QY 934 SerProIleGlyLeuAlaValAlaArgValThr 944  
DB 2043 AGCCCATTTGGGCTAGCCGTGGCCCGGGGTACA 2075

RESULT 14  
ID AAK83061 standard; DNA; 2077 BP.  
AC AAK83061;  
XX 07-NOV-2001 (first entry)  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37873.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KM Cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX MO200157182-A2.  
PD 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227109.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251899.  
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI, 2001-483426/52.  
 DR  
 XX  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 XX  
 XX  
 PS Disclosure; SEQ ID NO 37873; 3071bp + Sequence Listing; English.  
 CC  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (i)  
 CC amino acid sequences given in AAM62170 to AAM91921. (i) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (i)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (i) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (i) by expressing inactive proteins or to  
 CC supplement the patient's own production of (i). Additionally, (i)  
 CC polynucleotides may be used to produce the secreted (i), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (i) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK4703  
 CC to AAK7694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169  
 CC represent sequences used in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 2077 BP; 451 A; 586 C; 599 G; 441 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2077  
 Score: 691.00 Matches: 691  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.64% Indels: 0  
 DB: Gaps: 0  
 US-09-916-849a-3 (1-2923) x AAK83061 (1-2077)  
 QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
 DB 3 ACCCAAGCTTCAAGGCTCAGCGGAGGACCAAGGCAATGCCCAAGAGTGCCTGAGCT 62  
 QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlu 293  
 DB 63 AACACCAACCACTTGGTTACGACCAACCAATGACCACTGTTGAGCAGCAGAG 122  
 QY 294 TyrIleGluSerIleuAspGluAsnLeuGluValGlyTyrGluValLeuThrValArgAla 313  
 DB 123 TACAAAGAGACCTTCAGGAGAACTCGAGGTGGCTATGAGGTGCTCAGTCAGGACC 182  
 QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGly 333  
 DB 183 ACGAGTGTGATGCCCTCCCAATGCCAATATCTGTACCCGCTGAGAGGGGCTCGGG 242  
 QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
 DB 243 GGAGACCCCTCTGAAGTCTTTGAGATGACCCCTGCTGGGGATCCGAACCCGTGGC 302  
 QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
 DB 303 CCGTGTGATCGGAGAGGTGAAATCCACAGCTGAGGAGGCAAGTGAACCAAGGCT 362  
 QY 374 ArgAspProGlyProArgSerThrThrAlaIleValPheLeuSerValGluAspAspAsn 393  
 DB 363 CGGACCCGGGCTCTCGAGATCACACAGCCGCTGTTTCTTCTTCTGTGAGAGATGACAT 422

Qy	394	AepbanaIaPProGInPheSerGluYuaPrgYrVaIValGInVaIaPrgIuaPrgVaIaThr	413
Db	423	GATATATGCCCCCAAGTTTATGTGAGAAAGCGCTATGTGTGCTCAAGTGAAGGAAGATGTACT	482
Qy	414	ProGInVaIaProVaIleuApgVaIaThraIaSerAepaIaPrgAaPrgYglYSerAaMaIaVaI	433
Db	483	CCAGGGGGCCCAAGTACTCCGAGTCCACAGCTCGATTCAGACCAAGGGAGCAATCCCGTG	542
Qy	434	ValHaSTYrSerIleMeSerGluYaaMaIaHrArgYglInPheTYrLeuAaPrgIaGlnThr	453
Db	543	GTGCCTATAGATCATAGATGAGGCAATCTCGGGGACAGTTTATCTGAATGCCAGACT	602
Qy	454	GlyAlaIaAepVaIaIaSerProIeuaAaPYrGluThrThrIlyGluTYrThrLeuApg	473
Db	603	GGAGCTCTGATATGTGTAGAGCCCTCTTACATATAGAGACCAACAGATCACCTTAAG	662
Qy	474	ValATrgIaGInAaPrgIyGlyAaPrgProIeuaSerAaMaIaSerGlyLeuVaIaThraI	493
Db	663	GTGGAGAGCAAGAAAGTGGCCGCTCCCACTCTTAATGTCTCGGCTTGGTGAACATA	722
Qy	494	GInVaIleuAaPrgIleAaMaPrgMaIaProIlePheVaIaSerThrProPheGlnIaHr	513
Db	723	CAGGTCCCTGGAATTAAGACAAATCCCAATCTTCGTCAGAGACCCCTTTCAGGCTACT	782
Qy	514	ValleuGInuSerVaIaProIeugIYTYrLeuVaIleuHIsvalGlnaIaIleAaPrlaAaP	533
Db	783	GTCCCTGGAGAGTGTCCCTTAGGCTACTCGTGTCTCCATGTCCAGGCTATGACGCTGAT	842
Qy	534	AlaGlaYaPaaMaIaAaTgLeuGlnTYrAaPrgIeuaIaGlyVaIaGlyHIsAaPrgPhe	553
Db	843	GCTGGTGCAATGGCCCGCTGGAAATACCGCTGTGGGGTGAAGACATGACTTCCCTTC	902
Qy	554	ThrlIeaAaMaGlyTYrGlyTYrIleSerVaIaIaIaGluIeuaAaPrgIuGluVaI	573
Db	903	ACCATCAACAAATGGACAGAGCGTGAATCTGTGGCTGTGAATGAACTGACCGGGAAGATT	962
Qy	574	AaPheTYrSerPheGlyVaIaGlnIaAaPrgAaPrgIaGlyTYrProIaIeuaThraIaSer	593
Db	963	GATTTCACAGCTTTGGGGTGAACCTGAGACCAATGGACCTCCAGACTCACTGCCCTCG	1022
Qy	594	AlaSerVaIaSerVaIaThraIleuAaPrgVaIaAaMaPaaMaPrgThrPheThrGInPro	613
Db	1023	GCCAGTGTCAAGGTGACTGTCTCGATGTCAACGACAAACAACTTTATCCCAACA	1082
Qy	614	GluTYrThraIaAaPrgLeuAaMaIaAaIaValaGlyThrSerVaIaThraIaIaSer	633
Db	1083	GAGTACACAGTGGCGCTCAATAGAGATCAAGCTGTGGGACCAAGCGTGGTGAACGGTCA	1142
Qy	634	AlaVaIaAaPrgAaPrgAaHIsaSerVaIleThrTYrGlnIleThrSerGluYaaMaHrAaPrg	653
Db	1143	GCTGTGACCGGATGCTCATATGTGTCAATCACTACGAATACCAAGTGGCAATACTGA	1202
Qy	654	AaMaPrgPheSerIleThrSerGInuSerGlyGlyIleuVaIaSerIeuaIaAepProIeu	673
Db	1203	AACCGCTTCTCCATCACCAAGCAAAAGTGGTGGCTGTGATTCCTTGCCCTGCACATG	1262
Qy	674	AaPYrTYrVleuGlnAaPrgGlnTYrTYrVaIleuAaIaValaThraIaSerAaPrgIYThraGln	693
Db	1263	GACTACAAACTTGACCGGCAAGTATGTGGTGTGTACCGCTCCGAATGGCACTGGGAG	1322
Qy	694	AaPThraIaGlnIleVaIaIaMaIaThraAaPrlaAaMaThraHIsaAaPrgVaIaPheGln	713
Db	1323	GACAGGGCAAGATTGTGTGAATGTCAACCAAGCCCAACCACTGTCCTGTCTTTGAG	1382
Qy	714	SerSerHIsTYrThraIaMaIaMaIaMaIaAaPrgProIaIaGlyThrThraIaIleu	733
Db	1383	AGCTTCCCACTTAACAGTAAATGTATATAGGACCGCGCGGAGGACCAACGATGTGTGTG	1442
Qy	734	IleSerIaThraAaPrgIuaAaPrgThrGlyGlnaMaIaAaTYrIleThrTYrPheGInuAaP	753
Db	1443	ATCACGGCCACGGAATAGAGACACAGGTGAATGCCCGCATCACTTACTATGAGAGAC	1502
Qy	754	SerIleProGInPheAaPrgIleAaPrlaAaPrgThraIaIaIaIaThrThrGlnIaGluIeu	773

[illegible]





polynucleotides and proteins are used to prevent, treat a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For example, disorders associated with decreased expression of SPs. The SP polynucleotide or a vector expressing them may be administered to treat diseases by gene therapy. Antisense molecules may be administered to down regulate expression of SPs by binding with the cells own genes and preventing their expression. The polynucleotides may also be used as DNA probes in diagnostic assays. The SPs may also be used as antigens to produce antibodies and to identify modulators (agonists and antagonists) of the SPs. The anti-(SP) antibodies and antagonists may also be used to down regulate expression and activity of SP and as diagnostic agents for detecting the presence of SPs in samples. The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanoma), neoplasms of the breast or liver, Searay syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). Other uses include wound healing, maintenance of organs before transplantation, support of cell culture of primary tissues, modulation of for example

**Alignment Scores:**

Pred. No.:	0	Length:	2077
Score:	691.00	Matches:	691
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	23.64%	Indels:	0
AB:	22	Gaps:	0

US-09-916-849A-3 (1-2923) x AAS31491 (1-2077)

Qy	254	ThrHisValPheArgValThrlAglInasphIselGmetProArgArgSerAlaLeuAla	273
Db	2075	ACCACAGCTTCAGGGTCACGGCCGAGGACCAACGGCAGGCCCCACGAGAGGCCCTGGCT	2016
Qy	274	ThrLeuThrIleLeuValThrAspThzAsnAspHisAspProValPheGIGlnGlnu	293
Db	2015	ACACTCACCACTCTGGTTACTGACACCAATGACCACTGACCTGTGTTCGACGACGAGAG	1956
Qy	294	TyrIysGlnSerLeuArgIuAsnLeuGluValGlyYrGluValLeuThrValArgAla	313
Db	1955	TACAAGAGAGACCTCAGGGAGAACCTGGAGGTTGGCTATGAGGTGCTCACTGCAGGCC	1896
Qy	314	ThrAspGlyAspAlaProProAsnAlaAsnIleLeuYrYrAglLeuLeuGlySerGly	333
Db	1895	ACGATGTGTATGGCCCTCCCAATGCCAATATTCTGTACCCCTGCTGAGAGGGGTCTGGG	1836
Qy	334	GlySerProSerGluValPheGlnIleAspProArgSerGlyValIleArgThrArgGly	353
Db	1835	GGCAGCCCCCTGTGAAGTCTTGAAGTGCACCTCGCTGGGGGTGATCCGAACCCGTGGC	1776
Qy	354	ProValAspArgGluGluValGlnSerYrGlnLeuThrValGlnAlaSerAspGlnGly	373
Db	1775	CCTGTGATCGGGAGAGGTGGAACTCCACAGCTGACGGTGAAGGACCAATGACCAAGGT	1716
Qy	374	ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn	393
Db	1715	CGGAGACCGGGGTCTCTCGAGATACCAACAGCCCTGTTTCTCTTCTGTGAGAGATGACAT	1656
Qy	394	AspAsnAlaProGlnPheSerGlnIuAspArgYrValValGlnValArgGluAspValThr	413
Db	1655	GATATATGCCCCCAGGTTTATGTGAAGAGGCTATGTGTGTCACAGTGAAGGAGAGATGTACT	1596
Qy	414	ProGlyAlaProValLeuArgValThzAlaSerAspArgAspGlySerAsnAlaVal	433
Db	1595	CCAAGGGGCCCAAGTACTCCAGTCAACAGCTCGGATTCAGACACAGGGAGCAATGCCCTG	1536
Qy	434	ValHisYrSerIleMetSerGlyAsnAlaArgGlyGlnPheYrLeuAspAlaGlnThr	453
Db	1535	GTGACATATAGCATCATGATGGGCAATGCTCGGGGACAGTTTATCTGAGGCCAGACT	1476

OY	454	G1yAlaLeuAspValValSerProLeuAspTyrGluThrThrLysGluTyrThrLeuArg	473
Db	1475	GGAGCTCTGGATGTGGTGAACCTCTTGACTATGAGACGACCAAGAGTACACCTCACCG	141
OY	474	ValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrVal	493
Db	1415	GTGGAGGACAGGATGTGGCCGTCCTCCACTCTTAATGTCTTGGCTTGGTACAGTA	135
OY	494	GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr	513
Db	1355	CAGGCTCGATATCAACGAAATGCCCTCACTTCGTACGACACCCCTTTCAGGGTCACT	129
OY	514	ValLeuGluSerValProLeuGlyTyrTyrLeuValLeuHisValGlnAlaIleAspAlaAsp	533
Db	1295	GTCTCGAGAAATGTCCCTTAAAGCTACTGTGTCTTCAGATGTCCAGGCTATTCACGCTGAT	123
OY	534	AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe	553
Db	1235	GCTGGTGCATATGCCCGCTGGAAATACCGCTTGTCTGGGGTGGACATGACTTCCCTTC	117
OY	554	ThrIleAsnAsnGlyThrGlyTyrPheSerValAlaAlaGluLeuAspArgGluGluVal	573
Db	1175	ACCATCAACAATGGACACAGCTGGATCTCTGTGGCTGTCACTGACCGGAGGAGAGTT	111
OY	574	AspPheTyrSerPheGlyValGlnAlaArgAspHisGlyThrProAlaLeuThrAlaSer	593
Db	1115	GATTTCTACACTTTGGGGTGTGAAGCTGAGACCAATGGACATCCAGCACTCACTGCTCTG	105
OY	594	AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro	613
Db	1055	GCCAGTGCACCGGACGCTGCTCTGGATGTCAACGCAACAATTCACACTTTCACCAACA	996
OY	614	GluTyrThrValArgLeuAsnGlnAspAlaAlaValGlyThrSerValValThrValSer	633
Db	995	GAGTACACAGTGGCGCTCATAGAGATCCAGCTGTGGGACCAAGGTGGTGAACGGTGTCA	936
OY	634	AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg	653
Db	935	GCTGTGACCGGTATGTCTCATAGTGTCAATCACTACCAATCACCAAGTGGCAATCTCGA	876
OY	654	AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu	673
Db	875	AACCGCTTCTCAATCACCAAGCAAGGTGGTGGCTGTATCTCTTGGCTTGGCACATG	816
OY	674	AspTyrIleLeuGlnArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln	693
Db	815	GACTACCAACTTGAAGCGGCACTATGTGTGGCTGTATACCGCTCCGATGGCACTGGCAG	756
OY	694	AspThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisArgProValPheGln	713
Db	755	GACACGGGACAGATTTGGTGAATGTCAACGACCAACCCCACTGCTGTCTTTCAG	696
OY	714	SerSerHisTyrThrValAsnValAsnGlnAspArgProAlaGlyThrThrValValLeu	733
Db	695	AGCTTCCACTATACGTGAATGTTAATAGAGACCGGCGGACGACCAACGGTGGTCTG	636
OY	734	IleSerAlaThrAspGluAspThrGlyGlyLeuAsnAlaArgGlyIleThrTyrPheMetGluAsp	753
Db	635	ATCAGCGGCAACGATGAGGACACAGGTGAATGCGGCATCACTTACTTATGAGGAGC	576
OY	754	SerIleProGlnPheAspIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeu	773
Db	575	AGCATCCCCCATGTTCCGATGATGACACACGGGGGTGTGTACCAACCAAGGCTAAGCTG	516
OY	774	AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro	793
Db	515	GACTATGAGACCAAGGTGTTTACACCTTGGCATTAATGTCTCGGACAAATGGCAATCCC	456
OY	794	GlnIleSerSerAspThrThrTyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaPro	813
Db	455	CAGAAAGTCCGACACCACTTACTGTGAAGTCTGTGGAAACGAGTAAATGACAAATACCCCT	396
OY	814	GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer	833

```
Db      395 CAGTTCTTCGAGACTCTCTACCGAGGCACTGTTATGAGATGTGCCACCTTCACTAGC 336
Qy      834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853
Db      335 GTCTGCAAGATCTCAAGCACTGATCGTGAATTCGAGCTTAATGCGAGGCTTTCTACACC 276
Qy      854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal 873
Db      275 TTCCAAGAGGCGGACGATGAGACGGTGACTTTATTGTGAGTCCACGTCAAGGCATCGTG 216
Qy      874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla 893
Db      215 CGAAGCCTACGAGAGGCTGATCGAGAGAACGTGGCCAGTATGCTTGGCGGCATATGCA 156
Qy      894 ValAspLysGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913
Db      155 GTGACAAAGGGGATGCCCCAGCCCGCACCTATGGAAGTGACAGTCACTGTGTTGANT 96
Qy      914 ValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGluGluAsn 933
Db      95 GTGAATGACAAATCCCCCTGCTTTTGAGCAGATGAGTTGATGTGTGTGGAAGAGAAC 36
Qy      934 SerProIleGlyLeuAlaValAlaArgValThr 944
Db      35 AGCCCCATTGGGCTAGCCCGTGGCCCGGGTCAACA 3
```

Search completed: February 14, 2004, 11:59:03  
Job time : 2012 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 11:15:14 ; Search time 302 Seconds

(without alignments)  
4272.063 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 2923  
Sequence: 1 MRSRTGVPILPPPPPLLL.....AGTDESSGSEPLFFFLH 2923

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO/US0916849/runat.11022004.143918.21374/app.query.fasta\_1.3079  
-DB=Issued Patents NA -QFMT=fastexp -SUFFIX=Oligo.rn1 -MINMATCH=0.1 -LOOPTL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0916849 @CGN 1.1 169 @runat.11022004.143918.21374 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	0.7	135	1	US-07-998-003A-15
2	21	0.7	135	1	US-08-453-274B-15
3	21	0.7	135	1	US-08-453-695A-15
4	21	0.7	135	1	US-08-268-161A-15
5	21	0.7	135	2	US-08-453-702A-15
6	21	0.7	135	3	US-09-099-639-15
7	21	0.7	135	5	PCT-US93-12588-15
8	21	0.7	135	5	PCT-US95-08071-15
9	18	0.6	4566	2	US-08-465-976A-1
10	18	0.6	4566	2	US-08-982-412-1
11	15	0.5	131	1	US-07-998-003A-82
12	15	0.5	131	1	US-08-453-274B-82

13	15	0.5	131	1	US-08-453-695A-82	Sequence 82, Appl
14	15	0.5	131	1	US-08-268-161A-82	Sequence 82, Appl
15	15	0.5	131	1	US-08-453-702A-82	Sequence 82, Appl
16	15	0.5	131	3	US-09-099-639-82	Sequence 82, Appl
17	15	0.5	131	5	PCT-US93-12588-82	Sequence 82, Appl
18	15	0.5	131	5	PCT-US95-08071-82	Sequence 82, Appl
19	15	0.5	429	3	US-08-905-223-61	Sequence 61, Appl
20	14	0.5	699	3	US-09-276-531-91	Sequence 91, Appl
21	13	0.4	1362	2	US-08-318-837-8	Sequence 8, Appl
22	13	0.4	2023	4	US-09-491-522-6	Sequence 6, Appl
23	13	0.4	2450	4	US-09-491-522-2	Sequence 2, Appl
24	13	0.4	6692	4	US-09-491-522-1	Sequence 1, Appl
25	12	0.4	252	2	US-08-623-906A-1	Sequence 1, Appl
26	12	0.4	1011	4	US-08-811-481-27	Sequence 27, Appl
27	12	0.4	1103	3	US-08-927-219-54	Sequence 54, Appl
28	12	0.4	1946	2	US-08-861-464-1	Sequence 1, Appl
29	12	0.4	1946	2	US-08-396-001-1	Sequence 1, Appl
30	12	0.4	1946	3	US-09-323-433A-1	Sequence 1, Appl
31	12	0.4	2050	4	US-09-620-312D-761	Sequence 761, App
32	12	0.4	2426	3	US-09-028-312D-2	Sequence 2, Appl
33	12	0.4	2943	2	US-08-548-159-2	Sequence 2, Appl
34	12	0.4	2994	2	US-08-548-159-4	Sequence 4, Appl
35	12	0.4	3132	2	US-08-224-482-3	Sequence 3, Appl
36	12	0.4	3132	3	US-09-205-921-1	Sequence 1, Appl
37	12	0.4	3132	4	US-09-300-958A-32	Sequence 32, Appl
38	12	0.4	3132	4	US-09-702-705-320	Sequence 320, App
39	12	0.4	3132	4	US-09-736-457-320	Sequence 15, Appl
40	12	0.4	3287	4	US-08-811-481-15	Sequence 5, Appl
41	12	0.4	3311	4	US-09-367-891A-5	Sequence 1, Appl
42	12	0.4	580073	4	US-08-545-528D-1	Sequence 1, Appl
43	11	0.4	36	2	US-08-863-639A-31	Sequence 31, Appl
44	11	0.4	48	4	US-08-589-109A-11	Sequence 11, Appl
45	11	0.4	57	3	US-08-894-511-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-07-998-003A-15  
; Sequence 15, Application US/07998003A  
; Patent No. 5643781  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESS: Bicknell  
; STREET: 20 South Clark Street  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/998,003A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5643781and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 30903  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/346-5750  
; TELEFAX: 312/984-9740  
; TELE: 25-3856  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 135 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-07-998-003A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-07-998-003A-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
Db 70 CCACTTCCAAAGCTCCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
Db 130 CCC 132

RESULT 2  
US-08-453-274B-15  
Sequence 15, Application US/08453274B  
Patent No. 5663300

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borum  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-274B-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-274B-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
Db 70 CCACTTCCAAAGCTCCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
Db 130 CCC 132

RESULT 3  
US-08-453-695A-15  
Sequence 15, Application US/08453695A  
Patent No. 5708143

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,695A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-695A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-695A-15 (1-135)

QY 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
Db 70 CCACTTCCAAAGCTCCCGGTCTAGTAACCGTGCAGGTCTCTAGACATCAACGCAATGCC 129

QY 503 Pro 503  
Db 130 CCC 132

RESULT 4  
US-08-268-161A-15  
Sequence 15, Application US/08268161A  
Patent No. 5798224

GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,161A  
FILING DATE: June 27, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Young J. Sub  
REGISTRATION NUMBER: P-41,337  
REFERENCE/DOCKET NUMBER: 27866/32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-268-161A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-268-161A-15 (1-135)

QY 483 ProleuseranValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCTCCGCTCTAGTAACCGTGCAGGCTCTGACATCAACGACAAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 5  
US-08-453-702A-15  
Sequence 15, Application US/08453702A  
Patent No. 5891706  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,702A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5891706and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32657  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-453-702A-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-702A-15 (1-135)

QY 483 ProleuseranValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCTCCGCTCTAGTAACCGTGCAGGCTCTGACATCAACGACAAATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 6  
US-09-099-639-15  
Sequence 15, Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSER: Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161  
FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Greta E. No. 6262237and  
REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-099-639-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 3 Gaps: 0

US-09-916-849A-3 (1-2923) x US-09-099-639-15 (1-135)

QY 483 ProLeuserAnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCCGCGTCTAGTACCGTCGACGCTCCTAGACATCAACGACATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 7  
PCT-US93-12588-15  
Sequence 15, Application PC/TUS9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESS: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
CLASSIFICATION:  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
PCT-US93-12588-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.72% Indels: 0  
DB: 5 Gaps: 0

US-09-916-849A-3 (1-2923) x PCT-US93-12588-15 (1-135)

QY 483 ProLeuserAnValSerGlyLeuValThrValGlnValIleuAspIleAsnAspAsnAla 502  
DB 70 CCACCTTCCACGCTCCGCGTCTAGTACCGTCGACGCTCCTAGACATCAACGACATGCC 129

QY 503 Pro 503  
DB 130 CCC 132

RESULT 8  
PCT-US95-08071-15  
Sequence 15, Application PC/TUS9508071  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESS: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08071  
CLASSIFICATION:  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE: 23 DEC 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32149  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-08071-15

Alignment Scores:  
Pred. No.: 1.76e-10 Length: 135  
Score: 21.00 Matches: 21  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 0.72% Indels: 0  
DB: 5 Gaps: 0

US-09-916-849A-3 (1-2923) x PCT-US95-08071-15 (1-135)

Qy 483 ProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleLeuAspAsnAla 502  
Db 70 CCACTTCCACGCTTCGGCTAGTAACCGTACGAGCTCCTGACATCAACGACAAATGCC 129

Qy 503 Pro 503  
Db 130 CCC 132

RESULT 9  
US-08-465-976A-1  
; Sequence 1, Application US/08465976A  
; Patent No. 5869632  
; GENERAL INFORMATION:  
; APPLICANT: SOBET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: ROSEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARILLA, BYRNE, BAIN GILFILLAN, CECCHI  
; ADDRESSEE: STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,976A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY P  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-444  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 994-1744  
; TELEFAX: (201) 994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 212..2863  
; US-08-465-976A-1

Alignment Scores:  
Pred. No.: 3.09e-06 Length: 4566  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.62% Indels: 0  
DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-465-976A-1 (1-4566)

Qy 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTirp 2512  
Db 869 GGCCTTGCTGTGGGCTGAGCCCTGAGGCTATGGGAACCTGACTTCTGCTGG 922

RESULT 10  
US-08-982-412-1  
; Sequence 1, Application US/08982412  
; Patent No. 5958729  
; GENERAL INFORMATION:  
; APPLICANT: SOBET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: ROSEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE,  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/982,412  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROCKES, ANDERS A  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: P181PCT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 212..2863  
; US-08-982-412-1

Alignment Scores:  
Pred. No.: 3.09e-06 Length: 4566  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.62% Indels: 0  
DB: 2 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-982-412-1 (1-4566)

Qy 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTirp 2512  
Db 869 GGCCTTGCTGTGGGCTGAGCCCTGAGGCTATGGGAACCTGACTTCTGCTGG 922

RESULT 11  
US-07-998-003A-82  
; Sequence 82, Application US/07998003A  
; Patent No. 5643781  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESSEE: Bicknell  
; STREET: 20 South Clark Street



CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/998.003A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5643781and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 30903  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/346-5750  
TELEFAX: 312/984-9740  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-07-998-003A-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-07-998-003A-82 (1-131)

QY 791 G|Y|I|E|P|r|o|g|I|n|y|S|e|r|a|p|h|T|h|T|r|e|n|g|I|n|I|e|u|V|a|l 805  
Db 61 GGCAATCCCTCAAAATGACACTACTATTGGAATCTTAGTA 105

RESULT 12  
US-08-453-274B-82  
Sequence 82, Application US/08453274B  
Patent No. 5663300  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453.274B  
FILING DATE: 30-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5663300and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32660  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-274B-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
DB: 1 Gaps: 0

US-09-916-849A-3 (1-2923) x US-08-453-274B-82 (1-131)

QY 791 G|Y|I|E|P|r|o|g|I|n|y|S|e|r|a|p|h|T|h|T|r|e|n|g|I|n|I|e|u|V|a|l 805  
Db 61 GGCAATCCCTCAAAATGACACTACTATTGGAATCTTAGTA 105

RESULT 13  
US-08-453-695A-82  
Sequence 82, Application US/08453695A  
Patent No. 5708143  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray, & Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453.695A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5708143and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 32658  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 131 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-453-695A-82

Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0

DB: 1 Gaps: 0  
US-09-916-849A-3 (1-2923) x US-08-453-695A-82 (1-131)  
Qy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105  
RESULT 14  
US-08-268-161A-82  
; Sequence 82, Application US/08268161A  
; Patent No. 5798224  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/268.161A  
; FILING DATE: June 27, 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Young J. Suh  
; REGISTRATION NUMBER: P-41,337  
; REFERENCE/DOCKET NUMBER: 27866/32149  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 82:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-268-161A-82  
Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
Gaps: 1  
DB: 1  
US-09-916-849A-3 (1-2923) x US-08-268-161A-82 (1-131)  
Qy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105  
RESULT 15  
US-08-453-702A-82  
; Sequence 82, Application US/08453702A  
; Patent No. 5891706  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: Protocadherin Materials and Methods  
; NUMBER OF SEQUENCES: 115  
; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
; ADDRESSSEE: Borun  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.702A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5891706and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 32657  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 82:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 131 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-453-702A-82  
Alignment Scores:  
Pred. No.: 5.61e-05 Length: 131  
Score: 15.00 Matches: 15  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.51% Indels: 0  
Gaps: 0  
DB: 2  
US-09-916-849A-3 (1-2923) x US-08-453-702A-82 (1-131)  
Qy 791 GlyIleProGlnIySeraPThrThrTyrLeuGluIleVal 805  
Db 61 GGCAATCCCTCAAAAATCAGACCTACTATTGGAAATCTTAGTA 105  
Search completed: February 14, 2004, 21:16:08  
Job time : 322 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 18:30:55 ; Search time 1943 Seconds  
(without alignments)  
5541.564 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 2923  
Sequence: 1 MRSPTGVPLPPPPPLLL.....AGTVDESSGSEFLFENFLH 2923

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4891994

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-O=/cgn2\_1/USPTO.spool/US09916849/runat.11022004.143919.21449/app.query.fasta\_1.3079  
-DB=Published Applications NA -GMT=fastap -SUPFX=01igo.rmpb -MINMATCH=0.1  
-LOOPT=0 -LOOPTXT=0 -NITS=bits -START=1 -END=1 -MATRIX=01igo  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR.SCOR=quality -THR.MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USRR=US09916849 @CGN 1.1 1417 @runat.11022004.143919.21449  
-NCPU=6 -ICPU=3 -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBIOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	2923	100.0	8772	9	US-09-788-711A-3	Sequence 3, App1
2	2923	100.0	10531	15	US-10-225-567A-523	Sequence 523, App
3	2837	97.1	8871	1	US-09-788-711A-1	Sequence 1, App1
4	1103	37.7	24370	12	US-10-292-798-931	Sequence 931, App
5	776	26.5	2332	9	US-09-764-870-569	Sequence 569, App
6	776	26.5	2332	15	US-10-125-540-569	Sequence 569, App
7	729	24.9	2391	15	US-10-176-847-99	Sequence 99, App1
8	691	23.6	2077	9	US-09-764-870-570	Sequence 570, App
9	637	23.6	2077	15	US-10-125-540-570	Sequence 570, App
10	537	18.4	2603	9	US-09-843-856-1	Sequence 1, App1
11	281	9.6	2695	12	US-10-264-237-636	Sequence 636, App
12	281	9.6	4152	10	US-09-925-300-359	Sequence 359, App
13	217	7.4	652	9	US-09-764-870-571	Sequence 571, App
14	217	7.4	652	15	US-10-125-540-571	Sequence 571, App
15	131	4.5	509	9	US-09-764-870-568	Sequence 568, App
16	131	4.5	509	15	US-10-125-540-568	Sequence 568, App
17	123	4.2	478	11	US-09-918-995-912	Sequence 912, App
18	118	4.0	658	9	US-09-764-870-66	Sequence 66, App1
19	118	4.0	658	15	US-10-125-540-66	Sequence 66, App1
20	105	3.6	443	9	US-09-764-870-214	Sequence 214, App
21	105	3.6	443	15	US-10-125-540-214	Sequence 214, App
22	102	3.5	721	9	US-09-764-853-931	Sequence 931, App
23	102	3.5	721	9	US-09-764-853-932	Sequence 932, App
24	102	3.5	721	11	US-09-764-881-168	Sequence 168, App
25	102	3.5	721	11	US-09-764-881-169	Sequence 169, App
26	102	3.5	721	12	US-10-242-747-168	Sequence 168, App
27	102	3.5	721	12	US-10-242-747-169	Sequence 169, App
28	102	3.5	721	15	US-10-073-865-152	Sequence 152, App
29	102	3.5	721	15	US-10-073-865-153	Sequence 153, App
30	102	3.5	754	9	US-09-764-898-70	Sequence 70, App1
31	96	3.3	925	9	US-09-764-870-70	Sequence 70, App1
32	96	3.3	925	15	US-10-125-540-70	Sequence 70, App1
33	82	2.8	290	9	US-09-764-853-239	Sequence 239, App
34	82	2.8	290	9	US-09-764-898-142	Sequence 142, App
35	82	2.8	290	11	US-09-764-881-16	Sequence 16, App1
36	82	2.8	290	12	US-10-242-747-16	Sequence 16, App1
37	82	2.8	290	15	US-10-073-865-49	Sequence 49, App1
38	82	2.8	290	15	US-09-843-856-15	Sequence 15, App1
39	79	2.7	11002	13	US-10-017-161-1095	Sequence 1095, App
40	70	2.4	321	9	US-09-843-856-18	Sequence 18, App1
41	61	2.1	444	9	US-09-843-856-14	Sequence 14, App1
42	61	2.1	1856	10	US-09-764-864-661	Sequence 661, App
43	59	2.0	925	13	US-10-017-161-1687	Sequence 1687, App
44	55	1.9	2201	10	US-09-764-864-233	Sequence 233, App
45	53	1.8	300	9	US-09-843-856-19	Sequence 19, App1

## ALIGNMENTS

RESULT 1  
US-09-788-711A-3  
; Sequence 3, Application US/09788711A  
; Patent No. US20020058328A1  
GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Teeta  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788, 711A  
; PRIORITY FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 8772  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-3

Alignment Scores:  
Pred. No.: 0  
Score: 2923.00  
Percent Similarity: 100.00%

Length: 8772  
Matches: 2923  
Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x US-09-788-711a-3 (1-8772)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 1 ATCCGAGCCCGGCCACCGGCGTCCCCCTCCCAACCGCGCGCGTGGCTGCTG 60  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 61 TTGCTGCTCTGCTGGCGCGCCCACTATTGGAGAACCAAGTGGGCGCTGTGCTCTTG 120  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAcCysAlaProMetGlyThrLeuCysProSer 60  
 Db 121 GGGTCAGGGGAGAGAGGCTTTGGGGGGCTGGCCCCCAATGGGCTGTGCTGATCC 180  
 QY 61 SerAlaSerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 181 TAGCGCTCGAACCTTGCTCTACACAGCGCTGCAAGGGAATGCGGGCACTGAGTGA 240  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTTTCysProGluSerGlyAlaHis 100  
 Db 241 GGCCACTGTGATCCCAACCAAGATGGCTGAGGGTTGTGTCCAGAAATCCAGGGCCAT 300  
 QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
 Db 301 ATTCCCTTACACAGAGCTCTGAGAGCTGCTGGAGCTGTGCTCTTGGGCAATTGGA 360  
 QY 121 GlyHisLeuSerProGluGlyLeuLeuThrLeuProGluGlyIleAspCysLeuValAla 140  
 Db 361 GGCCACCTTTCCCAACAGGGGCAAGCTCACACTGCCGAGAGAGACCCGTTAAAGGCT 420  
 QY 141 ProArgLeuArgCysGluSerCysGlyLeuAlaGlnAlaProGlyLeuArgValGlyGlu 160  
 Db 421 CCACGGCTAGAGTCCAGTCTGCAAGCTGGCAAGGCCCCGGGTGAGGGAGGGGAA 480  
 QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgGlyArgAsnValAsnThrAlaProGln 180  
 Db 481 AGGTACCAAGAGAGTCCCTGGGTGGCGCTGGAAAGAAATGTAAATACAGCCCCCAG 540  
 QY 181 PheGlnProProSerGlyGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db 541 TTCCAGCCCCCGAGCTACAGAGCCAGAGTCCGGAACACAGCAGCAGCAGCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyValGlyValArgLeuGlyThrMet 220  
 Db 601 GCAATCCCTAGAGGCGCATTCACCCGAGCAGAGGTGAGGCAAGTCTGAGATGACCATG 660  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
 Db 661 GATGGCCCTTTTATAGCCGCTCCACAGTTCTTCTCTGAGCCCACTGAGTGA 720  
 QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrThrLeuSerThrHisValPheArgValThr 260  
 Db 721 GTAAACCAAGCCGAGAGCTGATCTGTAGAACCAAGACCAACCACTTCCAGGGTCAAG 780  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 781 GCGCAGAGACCAAGGAGTCCCGCAGAGAGGCTTGCTGACCTCACTCACTTGGTTACT 840  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGlyValGlyGluSerLeuArgGlu 300  
 Db 841 GACACCAATGACATGACCTGTGTTCGAGCAGCAGAGATGACAGAGAGCTCAGGGAG 900  
 QY 301 AsnLeuGlnValGlyThrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGGAGGTGGCTATAGAGTGTCACTGTCAAGGCCACGATGATGATGCTCTCC 960  
 QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGlyGlySerGlyGlySerProSerGlnValPhe 340  
 Db 961 AATGCCAATATTTCTGTATCCGCTGCTGAGGGGTCTGGGGGAGAGCCCTCTGAAAGTCTTT 1020

QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 1021 GAGATGACCTTCGCTCTGGGGTATATCCGAACCCGTGGCCCTGTGATCGGGAGAGGTG 1080  
 QY 361 GluSerThrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1081 GAATCTTACAGCTGACCGTGAAGGCAAGTACCAAGGGTGGGAGCCGGGTCTCTGAGT 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGluPheSer 400  
 Db 1141 ACCACAGCCGCTGTTTCTCTTGTGTGAGATGACAAATGATATGCCCCCACTTATAGT 1200  
 QY 401 GlyIleAspThrValIleGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db 1201 GAGAGGCCCTATGTGTCTCAAGTGAAGGAGATGTGACTCCAGGGGCCCACTACTCCGA 1260  
 QY 421 ValThrAlaSerAspAspArgAspGlySerAsnAlaValIleHisThrSerIleMetSer 440  
 Db 1261 GTCAAGCCCTCGATCGAGACAGAGGGCAATGCCGTGGTGCATATAGCATCATGAGT 1320  
 QY 441 GlyAsnAlaArgGlyGlnPheThrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 Db 1321 GGCAATGCTCGAGACAGTTTATCTGATGCCCAAGTGAAGCTGTGATGTGAGTGA 1380  
 QY 461 ProLeuAspThrGluThrThrLeuGlyValThrLeuArgValArgAlaGlnAspGlyGly 480  
 Db 1381 CCTGTGACTATGAGACGACCAAGAGATGACCTTACGGGTGCCAGACCAAGATGTGTC 1440  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1441 CGTCCCCCACTCTAATGTCTGTGGCTTGGTGAAGATGAGTCTTGATATCAACAC 1500  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 1501 AATGCCCATCTTCTGAGACCCCTTCCAGGCTACTGTGAGAGAGTCCCTTA 1560  
 QY 521 GlyThrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db 1561 GGCTACCTGTTCTCCAGTCCAGGCTATGACGCTGATGTGAGCAATGCCCGCTG 1620  
 QY 541 GluThrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAspGlyThrGly 560  
 Db 1621 GAAATACCGCTTGTGGGTGGACATGACTTCCCTTCAACATCAACAAATGGCAAGGC 1680  
 QY 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheThrSerPheGlyVal 580  
 Db 1681 TGGATCTGTGTGCTGCTGAATGGAACGGAGAGAAATGATTTCTACAGCTTGGGTA 1740  
 QY 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1741 GAACTCGAGACATAGGACTCCAGCACTGCTGCGCCAGGTGACGCTGACTGTC 1800  
 QY 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGlyThrValArgLeuAsn 620  
 Db 1801 CTGATGTCAACGACMAATCAACCTTATCCAAACCAAGATCAACAGTGGCTCAAT 1860  
 QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 Db 1861 GAGATGACCTGTGGGACCAAGCTGTGACGATGACCTGTGACCTGATGCTCAT 1920  
 QY 641 SerValIleThrThrGlnIleThrSerGlyValAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 1921 AGTGCATCACTTACAGATACCAAGTGGCAATCTCGAAACCGCTTCCATCAACCGC 1980  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspThrThrLeuGluGluGln 680  
 Db 1981 CAAGTGTGTGTGGCTGTATCCCTTGTCCATCGACATCAACAACTTGAAGGGGAG 2040  
 QY 681 ThrValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 2041 TATGTGTTGGCTGTATCCGCTCCGATGCACTGGGACAGACAGGCAATGTGTG 2100

QY AsnVal1ThrAspAlaAsnThrHisArgProVal1PheGlnSerSerHisGlyThrValAsn 720  
Db AATGACACCGAGCCCAACCCATGCTGCTTGTGAGAGCTCCACCTATACATGAT 2160  
QY ValaAsnGluAspArgProAlaGlyThrThrVal1ValLeu1LeuSerAlaThrAspGluAsp 740  
Db GTTATATGAGAACCGGCGCGGCAACGACGATGCTGATACAGCGCCACGATATAGAAC 2220  
QY ThrGlyGluAsnAlaArg1LeuThrTyrrPheMetGluAspSer1LeuProGlnPheArg1Le 760  
Db ACAGGTGAGAAATGCCGATCACTTACATGAGAGACAGCATCCCCAGTTCCGCAATC 2280  
QY AspAlaAspThrGlyAlaVal1ThrThrGln1AlaGluLeuAspTyrrGluAspGln1AlaSer 780  
Db GATGACAGACGCGGCGCTGTACACCCAGGCTGAGCTGAGCTACGAAAGACCAATGCT 2340  
QY TyrrThrLeuAla1LeuThrAlaArgAspAsnGly1LeuProGln1LeuSerAspThrTyrr 800  
Db TACACCTCGGCACTTACTGCTCGGCAATGGCAATCCCGAAGATCCGACACCACTTAC 2400  
QY LeuGlu1LeuVal1AsnAspVal1AsnAspAsnAlaProGlnPheLeuArgAspSerTyrr 820  
Db CTGGAGATCTCGGTGACAGCTGAATGACAAATGCCCTCAGTTCTCGAGACTCTTAC 2460  
QY GlnGlySerVal1TyrrGluAspVal1ProPheThrSerVal1LeuGln1LeuSerAlaThr 840  
Db CAGGCGCATGCTTATAGAGAAATGTGCACCTTCATGACGCTGCAATCTCAGCAATCT 2520  
QY AspArgAspSerGly1LeuAsnGly1ArgVal1PheTyrrThrPheGlnGly1GlyAspAspGly 860  
Db GATCGATGATTCGACCTTATATGACAGGCTCTTACACCTTCAGAGAGCGGACATGGA 2580  
QY AspGlyAspPhe1LeuAlaGluSerThrSerGly1LeuAlaGlyThrLeuArgArgLeuAsp 880  
Db GACCGTGACTTATGTTAGTGACGTCACGTCAGGCACTGCGCAACGCTACGAGGCTGAT 2640  
QY ArgGluAsnVal1AlaGlnTyrrVal1LeuArgAlaTyrrAlaValaAspGlyMetProPro 900  
Db CGAGAGAACGTGGCCCATGATGCTTGGCGGATATGACAGTGGACAGAGGATGCCCA 2700  
QY AlaArgThrProMetGluVal1ThrVal1ThrVal1LeuAspVal1AsnAspAsnProProVal 920  
Db GCCCGCACACTTATGAGAGTACACTGCTGTTGATGATGATGACAAATCCCCCTGTC 2760  
QY PheGluGluAspGluPheAspVal1PheVal1GluGluAsnSerPro1LeuAlaVal1 940  
Db TTTAGAGCAGATGATGATGATGTTTGTGAGAGAGAACGCCCATTTGGGCTAGCCGTG 2820  
QY AlaArgVal1ThrAlaThrAspProAspGluGlyThrAsnAlaGln1LeuMetTyrrGln1Le 960  
Db GCCCGGTGACAGCCACTGACCCCGATGAGCACCAATGCCAGATTATGTACCAGAT 2880  
QY Val1GluGlyAsn1LeuProGlyVal1PheGlnLeuAsp1LeuPheSerGly1GluThrAla 980  
Db GTGAGAGGCAACATCCCTGAGGCTCTTCCAGCTGACATCTTCTCCGGGAGACTGACACC 2940  
QY LeuValaAspLeuAspTyrrGluAspArgProGlyTyrrVal1LeuVal11LeuAlaThrSer 1000  
Db CTGGTAGACTTAGACTAGAGGACCGGCTGATGCTGCTGATCATCAGGCCACGTC 3000  
QY AlaProLeuVal1SerArgAla1ThrVal1HisValaArgLeuLeuAspArgAsnAspPro 1020  
Db GCTCCTCTGAGAGCGCGGCTACAGTCCAGCTCCCTCTTGCACCGAAAGAACCA 3060  
QY ProVal1LeuGlyAsnPheGlu1LeuPheAsnAsnTyrrVal1ThrAsnArgSerSer 1040  
Db CCAAGCTCGGCACTTTGAGATCTTTTCAACAACATGTCTCAACATCGCTCAAGCAGC 3120  
QY PheProGlyGlyAla1LeuArgVal1ProAlaHisAspProAsp1LeuSerAspSer1Leu 1060  
Db TTCCCTGGGGGTGCTGATGCGCGAGTACCTCCCATGACCTCGATATCTCAGATGCTG 3180  
QY ThrTyrrSerPheGluArgGly1AsnGluLeuSer1LeuVal1LeuAsnAlaSerThrGly 1080

Db ACTTACAGCTTTGAGCGGGAATGAATCACTGCTGCTGCTCAATGCTCCACCGGT 3240  
QY G1uLeuTy1LeuSerArgAlaLeuAspAsnAsnArgProLeuGluAla1LeuMetSerVal 1100  
Db GAGCTGAAGCTTACCCCGGCACTGACAAACACCGGCTCTGGAGGCACTATAGCTG 3300  
QY LeuVal1SerAspGlyVal1HisSerVal1ThrAlaGlnCysAlaLeuArgVal1Thr1Le1Le 1120  
Db CTGGGTGACAGCGCGTACACAGCTGACCGCCAGTGGCGGCTGCGTGTACCATATC 3360  
QY ThrAspGluMetLeuThrHisSer1LeuThrLeuArgLeuGluAspMetSerProGluArg 1140  
Db ACCATAGAGATGCTCACCCACAGATCACTGTGGCTGAGAGACATGTACCCGAGGCG 3420  
QY PheLeuSerProLeuLeuGly1LeuPhe1LeuAlaVal1Ala1LeuThrVal1LeuPro 1160  
Db TTCTCTGACACACTGCTAGGCTCTTATCTCAGGCGGTGGCCGCACTGCGCACCGCA 3480  
QY ProAspHisVal1Val1PheAsnVal1GlnArgAspThrAspAlaProGlyGlyHis1Le 1180  
Db CCGAGCACCGTGGTGTCTTCAAGTACAGGGGAGACCGACGCCCGGGGCGCACATC 3540  
QY LeuAsnVal1SerLeuSerVal1GlyGlnProProGly1ProGly1ProProPheLeu 1200  
Db CTCAACGTGAGCTGTGGTGGGCGACGCCAGGGGCCCGGGGCGCGCCCTTCTG 3600  
QY ProSerGluAspLeuGlnGluArgLeuTyrrLeuAsnArgSerLeuLeuThrAla1LeuSer 1220  
Db CCTCTAGAGACTCGAGAGGCGCTTATCTCAACCGCAAGCTGCTGACGCGCATCTCG 3660  
QY AlaGlnArgVal1LeuProPheAspAspAsn1LeuCysLeuArgGluProCysGluAsnTyrr 1240  
Db GCACAGCGCGGTGCTCTTCAACGACCAATCTGCTTGGGAGGCCCTGGAGAACTTAC 3720  
QY MetArgCysVal1SerVal1LeuArgPheAspSerSerAlaProPhe1Leu1LeuSerSer 1260  
Db ATGGCTGCGTGTGGTGGTGGCTGACTCTTCCGCGCTTCATCGGCTCTCTCTCC 3780  
QY Val1LeuPheArgPro1Leu1SerProVal1Gly1LeuArgCysArgCysProProGly1Phe 1280  
Db GTGCTCTTCCGCGCCATCCACCCCGTGGAGGGCTGGCTGCGCTCCCGCCCGGCTTC 3840  
QY ThrGlyAspTyrrCysGluThrGluVal1AspLeuCysTyrrSerArgProCysGlyProHis 1300  
Db ACCGGTGACTAGTGGAGACCGAGGTGACCTGCTGACTCGCGGCTGTGGGCCCCAC 3900  
QY GlyArgCysArgSerArgGluGlyTyrrThrCysLeuCysArgAspGlyTyrrThrGly 1320  
Db GGGGCTGCGGAGCGCGGAGGCGGCTTACACTGCTCTGTGTATGCTTACACAGGCT 3960  
QY G1uHisCysGluVal1SerAlaArgSerGlyArgCysThrProGlyVal1CysLysAsnGly 1340  
Db GAGCATGTATAGGTGATGCTGCTCTGAGGCGCTTGCACCCCGGTGTCTGCAAGATGCG 4020  
QY GlyThrCysVal1AsnLeuLeuVal1GlyGlyPheLysCysAspCysProSerGlyAspPhe 1360  
Db GGCACTGTGTCAACCTGCTGGTGGGCGTTTCAAGTGCATTTGCGCATCGAGACTTC 4080  
QY G1uLysProTyrrCysGlnVal1ThrThrArgSerPheProAlaHisSerPhe1LeuThrPhe 1380  
Db GAGAGGCTTACTGACAGGAGGACAGCGGAGCTTCCCGCCCACTCTTATCATCACTTT 4140  
QY ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
Db CGGGGCTGCGCAAGGCTTTCACTTCACTTCCGCTCTCGTTTGCACAAAGAGAGCGC 4200  
QY AspGlyLeuLeuLeuTyrrAsnGlyArgPheAsnGluLysHisAspPheVal1AlaGlu 1420  
Db GACGGGTGTGTGTATGATGAGGCGTTTCAATGAGAGATATGATTTGGGCGCTCGAG 4260  
QY Val11LeuGlnGluGlnVal1GluLeuThrPheSerAlaGlyGluSerThrThrValSer 1440

Db 4261 GTGATCCAGAGAGAGGTCCAGCTCACTTCTCTGACAGGGAGTCAACCAACGAGTGTCC 4320  
Qy 1441 ProPheValProGlyValSerAspGlyValTTPH1sthrValGlnLeuValTyr 1460  
Db 4321 CCAATGGTCCCGAGAGAGTCACTGATGSCCATGGGCAATACGGTGCAGCTGAAATCTAC 4380  
Qy 1461 AsnLysProLeuLeuGlyValInthGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
Db 4381 AATAAACCACTGTTGGGTGAGACAGAGGCTCCCAAGGGCCATCAGAGCAGAGAGTGGCT 4440  
Qy 1481 ValValThrValAlaArgGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4441 GTGTGACCGGTGATGCTGTGACACAGAGTGGCTTGGCTTGGATCTGCTGGGGC 4500  
Qy 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysSerLeuAspLeuThr 1520  
Db 4501 AACTTATCTCTGTCTGTCCCAAGGCAACCAAGGTGGCAGCAAGAAAGTCTTGAAATCTGACG 4560  
Qy 1521 GlyProLeuLeuLeuGlyValProAspLeuProGlyLeuSerPheProValArgMetArg 1540  
Db 4561 GGGCCCTGCTACTAGGCGGGGTGGCTGACCTGCGCCAGAGAGCTTCCAGTCCGAAATGGG 4620  
Qy 1541 GlnPheValGlyCysMetArgAspLeuGlnValAspSerArgH1stIleAspMetAlaAsp 1560  
Db 4621 CAGTTCGTGGCTGACATGCGGAACTGCAAGTGGACAGCGGCAATAGACATGGCTGAC 4680  
Qy 1561 PheAlaAlaAsnGlnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
Db 4681 TTCAATGCCCAACATGGCAACCGTCCCTGGCTCCCTCCAAAGAAAGCGTGTGACAGC 4740  
Qy 1581 AsnThrCysH1saanglyGlyThrCysValAsnGlnTTPAspAlaPheSerCysGluCys 1600  
Db 4741 AACACTTGGCAATGGGGGCACTGGCGTGAACCAAGTGGAGCCGTTCACTGACGTGCG 4800  
Qy 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlyMetAlaAsnProGlnH1sPheLeu 1620  
Db 4801 CCCCCTGGCTTGTGGGCAAGCTGGCCCAAGAAATGGCCAAATCCACAGCACTTCTCG 4860  
Qy 1621 GlySerSerLeuValAlaTTPH1sglyLeuSerLeuProIleSerGlnProTTPYLeu 1640  
Db 4861 GGCAGAGCGCTGGTGGCTGGCAATGGCTCTCGCTGCCATCTCCCAACCTGGTAACTC 4920  
Qy 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCTCAATGTTCCGACAGCCAGCCAGCGAGTGTCTGTGACAGGCAATCACAGGGGG 4980  
Qy 1661 ArgSerThrIleThrLeuGlnLeuArgGlyValH1sValMetLeuSerValGlyGlyThr 1680  
Db 4981 CGAGCAACATCACTCAAGCTHCAAGAGGGCACTGATGCTGACGCTGAGAGGGACA 5040  
Qy 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyArgAlaAsnAspGlyAspTTPH1s 1700  
Db 5041 GGGCTTCAAGGCTCTCTCTCCGTCTGAGCCAGGCGGGCCAAATAGCGTGACTGGCAC 5100  
Qy 1701 H1sAlaGlnLeuAlaLeuGlyAlaSerGlyValProGlyH1sAlaIleLeuSerPheAsp 1720  
Db 5101 CAGTCAAGAGTGGACATGGAGGCGAGGGGGGCTGGCCATGCAATTCGTCTCTTCAT 5160  
Qy 1721 TyrGlyGlnGlnArgAlaGlyValAsnLeuGlyProArgLeuH1sglyLeuH1sLeuSer 1740  
Db 5161 TATGGGCAAGAGAGAGAGAGGGGCACTGGGCCCCGGCTGCATGATCTGCACTGAGC 5220  
Qy 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAlaArgIlyPheArgGly 1760  
Db 5221 AACCTAACATGGGCGGAATACTGGGCGAGCGCGGTGGCGCTTGGCTTGGGGGC 5280  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer 1780  
Db 5281 TGTGTGAGGGGTGGTGGGTGAGGCAATACGCAAGAGGGGGTTPAACACCTCGAATCCAGC 5340  
Qy 1781 H1sGlyGlyLeuSerIleAsnValGlyGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGAGAGCAATCAAGTGAACAAAGGCTGTAGCTGCTGACCTTGTGACTCAAC 5400

Qy 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTTPAspSerTyrSerCysSerCysAsp 1820  
Db 5401 CCGTGTCCGCTTACAGCTAATGGCAACAGCACTGGAGACAGCTAATCTCGACGTGTAT 5460  
Qy 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlyH1sgln 1840  
Db 5461 CAGGTTACTATGATGACAACTGTAATGTGTGACCTGAACCCGTTGAGACCAAG 5520  
Qy 1841 SerValCysThrArgLysProSerAlaProH1sglyTyrThrCysGlyLysProProAsn 1860  
Db 5521 TCTGTGTACCCGCAAGAGCCAGTGTCCCAATGGCTAATACCTGCAAGTGTCCCAAT 5580  
Qy 1861 TyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysProArgGlyTTPTPGly 1880  
Db 5581 TACCTTGGGCAATCACTGATGACAGAGANTTACAGAGCTTGTCCCGTGGCTGGTGGGA 5640  
Qy 1881 H1sProThrCysGlyProCysAsnCysAspValSerIlyPheAspProAspCysAsn 1900  
Db 5641 CATCCCAATGTGGGCCAATGCAACTGTGATGTCAAGCAAGGCTTGAACCAAGCTGCAAC 5700  
Qy 1901 LysThrSerGlyGlyCysH1sGlyCysLysGlyAsnH1sTyrArgProProGlyLysProThr 1920  
Db 5701 AAGACAGCGGAGAGTGCATGCAAGAGAAACACTACCGGCCCCAGGCAAGCCACCC 5760  
Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyAsp 1940  
Db 5761 TGCCTCTTGTGATGCTGATCCCAAGGCTCTCTTGTCCAGAGTGTGTACCTGAGGAT 5820  
Qy 1941 GlyLysCysProCysLysProGlyValIleGlyValGlnCysAspArgCysAspAsnPro 1960  
Db 5821 GGCAGGTGTCCATGGAAGCAGAGTGTCAATGGGCGCTGAGTGAACCGCTGTGAACAACCT 5880  
Qy 1961 PheAlaGlnValThrThrsAsnGlyCysGlyValAsnTyrAspSerCysAspProArgAlaIle 1980  
Db 5881 TTTCGTAGGTACCAACCAATGCTGTGAAGTAATTAATGACAGCTGCCCAAGAGGAT 5940  
Qy 1981 GlnAlaGlyIleTTPTPProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
Db 5941 GAGGCTGGGATCTGGTGGGCCCCGATCCGCTTCCGAGCTGCTGCTCTCTGCTGCC 6000  
Qy 2001 LysGlySerPheGlyThrAlaValArgH1sCysAspGlyH1sArgGlyTTPLeuProPro 2020  
Db 6001 AAAAGCTTTTGGAGCTGTGTGGCACTGTGATGACACAGAGGGGTGGCTCCCCCA 6060  
Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLysGlyPheAlaGlyArgLeu 2040  
Db 6061 AACCTTCAATGCACTGACGTCATCACTTCTGAGAACTGAAGGGCTTGGCTGAGGGCTA 6120  
Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg 2060  
Db 6121 CAGCGGAATGATGCAAGGCTTACACTCAAGGGCCCTCCACAGCACTAGCCCTGCTCCGCC 6180  
Qy 2061 AsnAlaThrGlnH1sthrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
Db 6181 AAGCGCACAGACACACACTGCTTACCTTGGAGAGAGTCAAGGTGCTTACCAAGCTG 6240  
Qy 2081 AlaThrArgLeuLeuAlaH1sglyLeuSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6241 GCCAGCGGCTGTGGCCCAAGAGAGCAACCAAGCGGGGCTTTGGGTGTCTGCAACAG 6300  
Qy 2101 AspValH1sPheThrGlyAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6301 GAGGTGCACCTTCACTGAAGATCTGCTGGCGGTGGGAGGCGCTCTGTGACACAGCAAC 6360  
Qy 2121 LysArgH1stTPGlyLeuLeuIleGlnGlnThrGlyGlyTyrAlaIleThrLeuGlnH1s 2140  
Db 6361 AAGCGCACTGGAGGTGATTCAGAGACAGAGAGGGTGGCAACCGCTGCTCCAGCAC 6420  
Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgH1sthrTyrLeuSerProPhe 2160  
Db 6421 TATGAGGCTTACGCAAGTCCCTGGGCCAGAACTGGGCAACCTAAGCCCTTTC 6480

QY 2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspIysGlyAsnPheAla 2180  
Db 6481 ACCATGTCACCCCAACATTGTCACTTCGTAAGTGGCTGACCAAGGAACTTGTGCT 6540  
QY 2181 GlyAlaIysLeuProArgTyrGluAlaLeuArgGlyGluGlnProProAsnIleuThr 2200  
Db 6541 GGGGCCAAGCTCCCGCTACAGAGCCCTGCTGGGAGACAGCCCGGACCTTGAGACA 6600  
QY 2201 ThrValIleLeuProGlySerValPheArgGluThrProProValValArgProAlaGly 2220  
Db 6601 ACAGTCATCTGCTGAGTCTGCTTCACAGAGAGCCCGCTGGTCAAGGCCCGAGGC 6660  
QY 2221 ProGlyGluAlaGlnGluProGlyGluLeuAlaArgArgGlnArgArgHisProGlyLeu 2240  
Db 6661 CCGGAGAGGCCCAAGAGCAAGAGAGCTGGACCGGCAAGCGAGCCCGGAGCTG 6720  
QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
Db 6721 AGCCAGGCTGAGGCTGTGGCCAGCTCATCTACCGACCCCTGGCCGAGCTACTGCT 6780  
QY 2261 HisAsnTyrAspProAspIysArgSerLeuArgValProIysArgProIleIleAsnThr 2280  
Db 6781 CATACCTATGACCTGACAGAGCGAGCTTGAGAGTCCCAAGCCCATCATCAACA 6840  
QY 2281 ProValIleSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspIys 2300  
Db 6841 CCGGAGGAGCATCAGCGCTCATGATGATGAGAGCTTCTGCCCCGGGCTCTGACAA 6900  
QY 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleGlyVal 2320  
Db 6901 CCGGTCAAGGTCACTTCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6960  
QY 2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrSerAlaArgIleCysGlu 2340  
Db 6961 TTCTGGAACCAATTCATCTGTGTCAGTGGCAGAGGTGGTGGCCAGAGGCTGTGA 7020  
QY 2341 ValValIlePheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
Db 7021 GTGCGCTTCCGCAAG 7080  
QY 2361 ValLeuMetAspValSerArgArgGluAsnGlyIleLeuProLeuIleThrLeuThr 2380  
Db 7081 GTGCTCAAG 7140  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400  
Db 7141 TACGGGCTCTAGAGGTCACTGTGCTGCTTGTCTGCTCACTTCTTCTCTCACTCTC 7200  
QY 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
Db 7201 TTGCGTATCTGGCTCCCAACCAACAGCGCATCCGATCACTGACAGCTGCCCTGGGC 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
Db 7261 CTGGCTCAGCTGTCTTCTCTCTGGGAATCAACAGGCTGACCTCTTGTGCTGACA 7320  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTyrAlaLeuLeuGlu 2460  
Db 7321 GTCATTTGCCATCTGCTGCACTTCGTACCTGCACTTCTTCTGGGCTGTGTGAG 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
Db 7381 GCCCTTGACCTGTACCGGCACTCACTGAGGTGCCGATGTCACACCGGCCCCCAATGGCC 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db 7441 TTCTACTACAGCTGGGCTGGGAGCTGCTGCTTCACTACAGGCTTACCGGTGGAGCTG 7500  
QY 2501 AspProGluGlyTyrGlyAsnProAspPheCysTyrLeuSerIleTyrAspThrLeuIle 2520  
Db 7501 GACCCCGAGGGCTACCGGAACTCTGCTTGTGCTCTCATCATGACAGCGCTCATC 7560  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540

Db 7561 TGAAGTTTGGCTGGGCCCGGTGGCTTTGCTGCTGAGAGAGTCTTCTGACATCTG 7620  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyIleValGlyGlyProAlaSer 2560  
Db 7621 GGGGCCCGGCTCTGTGTGCTGCCAGCGGCAAGGCTTGTAGAAAGATCTTCTCTCG 7680  
QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeu 2580  
Db 7681 GGCTTGAGCCCTCTTGTGCGGTCTCTGCTGCTGAGCGCAGCTGTGCTGCACTG 7740  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrIlePheAlaAlaThrCysAsnCysIle 2600  
Db 7741 CTCTCTGCAACAGGAGACACCTCTCTTCCACTCACTTGTGCTACCTGCAATTGCAATC 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerIysGluValArgIysAlaLeu 2620  
Db 7801 CAGGCCCCCTTCACTTCTCTCTATGTGTGTCTTACCAAGAGGTCCGAAAGCACTC 7860  
QY 2621 LysLeuAlaCysSerArgIysProSerProAspProAlaLeuThrThrIleSerThrLeu 2640  
Db 7861 AAGCTTGCTGCAAGCCGCAAGCCCAAGCCCTGACCTGCTGACCAACAGTCCACCTG 7920  
QY 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db 7921 ACCTGTCTCTCAACTGCCCCAGCCCTACAGCAGATGGCGGCTGTACCAAGCTTACGGA 7980  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyIysSerGlnProSerTyrIle 2680  
Db 7981 GACTGGCCGCTCTCTGCAAGCAAGTGTCTGGGCAAGAGAGAGAGAGAGAGAGAGAG 8040  
QY 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db 8041 CCTTCTTGTGAGAGAGAGAGTCCGCACTGAACCTTGGCCAAAGGCCCCCTGGCTGGG 8100  
QY 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8101 GATCAGAGAGCCCTGTCTCTGGAAGTCAAGACAGAGAGATGATCTCGACAGCACTCC 8160  
QY 2721 AspSerAspLeuSerLeuGluAspAspGlnSerIleSerTyrAlaSerThrHisSerSer 2740  
Db 8161 GACATGACTGTCTTATGAAGACAGACAGAGTGTCTTATGTCTTATCCCACTCAACA 8220  
QY 2741 AspSerGluGluGluGluGluGluGluGluGluGluAlaAlaPheProGlyGluGlnGly 2760  
Db 8221 GACAGTGAAGAGAAAG 8280  
QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGluArgLeuProLeuHisSerThrProIysAsp 2780  
Db 8281 TGGGATAGCTGTGGGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 8340  
QY 2781 GlyGlyProGlyProGlyIysAlaProTyrProGlyIysAspPheGlyThrThrAlaGlyGlu 2800  
Db 8341 GGGGGCCCCAGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8400  
QY 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
Db 8401 AGTAGTGGCAACGGAGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8460  
QY 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisIleGlyIleLeuIys 2840  
Db 8461 GAGGGGTCCCTAGAGCCCTTCCAGGCTCTTCTCCAGAGAGAGAGAGAGAGAGAGAG 8500  
QY 2841 LysIysCysLeuProThrIleSerGlyIysSerSerLeuLeuArgLeuProLeuGlnGln 2860  
Db 8521 AAGAGATGTCTGCAACATCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8580  
QY 2861 CysThrGlySerSerArgIysSerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
Db 8581 TGCAAGAGGTCTTCCGGGGCTCTCTCCGCTAGTAGAGGAGAGAGAGAGAGAGAGAG 8640  
QY 2881 ArgProProProAlaGlnSerIleGlnGlnGlnLeuAsnGlyValMetProIleAlaMet 2900



Db 8641 CGCCACCGCCCGGAGAGCTCCAGAGACGTAGACGGGCTCATGCCCATCGCCATG 8700  
 QY 2901 SerIlaYalagIyThrValaApGuaApSerSerGlySerGluPheUleuPheAen 2920  
 Db 8701 AGCATCAAGCGACGACGAGTGAGTGAAGACTCGTCAAGCTCCGAAATTCCTCTTAAAC 8760  
 QY 2921 PheUleuHis 2923  
 Db 8761 TTCCTGCAAT 8769

## RESULT 2

US-10-225-567A-523  
 ; Sequence 523, Application US/10225567A  
 ; Publication No. US20030113798A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LifeSpan Biosciences  
 ; APPLICANT: Brown, Joseph P.  
 ; APPLICANT: Burner, Glena C.  
 ; APPLICANT: Roush, Christine L.  
 ; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
 ; FILE REFERENCE: 1920-4-4  
 ; CURRENT APPLICATION NUMBER: US/10/225,567A  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: 60/257,144  
 ; PRIOR FILING DATE: 2000-12-19  
 ; NUMBER OF SEQ ID NOS: 2292  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 523  
 ; LENGTH: 10531  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-225-567A-523

## Alignment Scores:

Pred. No.: 0 Length: 10531  
 Score: 2923.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 15 Gaps: 0

US-09-916-849A-3 (1-2923) x US-10-225-567A-523 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 63 ATCGGAGCGCGGCGACCGGCGTCCCTCCCAACGCGCGCGCGCGTGTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyValAspGlnValGlyProCysArgSerLeu 40  
 Db 123 TTGCTGCTGCTGCTGCGCGCGCAATTGGAGAACCAAGTGGGGCCCTGCTCTCTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyThrLeuCysProSer 60  
 Db 183 GGGTCCAGGGGAGCAAGGCTCTTCGGGGGCTGCGCCCCCATGGGCTGAGCTCTGTCATCC 242  
 QY 61 SerAlaSerAsnLeuThrLeuLeuYThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 243 TCAGCGCTGAACCTTGCTGCTACACCAAGCGCTGCGGAGTGGGGCACTGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisAspArgGlyLeuArgValTrrCysProGluSerGluHis 100  
 Db 303 GGGCACTGGTACCCCAACGAGTGGCTGAGGGTTGGTCCAGAAATCCAGAGGCCAT 362  
 QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyIleGly 120  
 Db 363 ATTCCTCAACACAGCTCTCTGAAGGCTGCGCCCTGAGAGCTGCTGCTCTGGGCAATTGGA 422  
 QY 121 GlyHisLeuSerProGluGlyValLeuThrLeuProGluGluHisProCysLeuVala 140  
 Db 423 GGGCACTTTTCCCAAGGAGGAGCTCACTGCCCAAGAGGACCCGCTTAAAGGCT 482  
 QY 141 ProArgLeuArgCysGlnSerCysValLeuAlaGlnAlaProGluLeuArgIleGlu 160

Db 483 CCAGGCTCAGATGACAGTGCATGCTGCAAGCTGACAGAGCCCGGCGCTCAGGAGGGAGA 542  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgValAsnThrAlaProGln 180  
 Db 543 AGTCACCAAGAAAGTCTCCGAGTGGCGCTCGAAAAGAAATGTAATACACCCCCCAG 602  
 QY 181 PheGlnProProSerSerGlyAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db 603 TTCAGACCCCAAGCTACACAGGCGCACAGTCCGAGAAACAGACGACGAGGACCCCTGTT 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyValaGlyValLeuGluTyrThrMet 220  
 Db 663 GCATCCCTGAGGGCCATGCAACCGGACGAGGTGAGGAGGTGCACTGAGTACACCAAG 722  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
 Db 723 GATGCCCTTTGATAGCGCGCTCCCAACAGTTCCTCTCCGAGACCACTGATGCA 782  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrIleSerThrHisValPheArgValThr 260  
 Db 783 GTAACCAAGCGCGAGAGGTGATGCTGAGACCAAGACCAACCACTTCAGGGTCAAG 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 843 GCGCAGAGCACCGGATGCCCGACGAAAGTGCCTGCTACACTCACCATTTGCTTACT 902  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrIleValSerLeuArgGlu 300  
 Db 903 GACACCAATGACCAATGACCTCTGTGTTCCAGACAGAGATACAGAGAGCTTCAGGGAG 962  
 QY 301 AsnLeuGluValaGlyTyrGluValLeuThrValaArgAlaThrAspGlyAspAlaProPro 320  
 Db 963 AACCTGAGAGTGGCTGATAGAGGTGCTCACTGTCAGGGCCAGCGATGTGTATCCCTCC 1022  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyValProSerProSerGluValPhe 340  
 Db 1023 AATGCCAATATTCGTATACCGCTGTGTGAGGGGCTTGAGGGGAGCCCTGTAAAGCTTT 1082  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 1083 GAGATCGAACCTCGCTGCGGTGATCCGAACCCCTGCGCTGTGATCGGAAAGAGTG 1142  
 QY 361 GluSerTyrGlnLeuThrValaGluAlaSerAspGlnIleArgAspProGlyProArgSer 380  
 Db 1143 GAATCCTACCACTGACGCTGAGGCAAGTGAACAGGGTCCGAGACCCGGGCTCTCGAGAT 1202  
 QY 381 ThrThrAlaAlaValPheLeuSerValaGluAspAsnAspAsnAlaProGlnPheSer 400  
 Db 1203 ACCACAGCGCGTGTTCCTTCTGTGAGAGTGAACAATGATTAATGCCCCCAATTTAGT 1262  
 QY 401 GluIleArgTyrValaValaGluValaArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db 1263 GAGAAAGCGTATGTGTCCAGGTGAGGAGATGTGATCCAGGGGCCCAATATCTCCGA 1322  
 QY 421 ValThrAlaSerAspArgAspIleGlySerAsnAlaValaHisTyrSerIleMetSer 440  
 Db 1323 GTACACGCTCCGATCGAGACCAAGGGAGCAATGCGTGTGCTACCTATGCACTGAGT 1382  
 QY 441 GlyAsnAlaArgGlyGluPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValaIleSer 460  
 Db 1383 GGCATGTCTGGGAGACGTTTATCTGATGTCCAGATCCGAGCTTGATGTGTGAGC 1442  
 QY 461 ProLeuAspTyrGluThrThrIleArgGluTyrThrLeuArgValaArgAlaGlnAspGly 480  
 Db 1443 CCTCTTGAATAGAGACGACCAAGAGATACCTTAACGGGTGCGAGCACAGATGTGTGAC 1502  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 1503 CGTCCCACTCTCTTAATTCCTGTGCTTGTGAGAGATGACAGGTCTGTAAATCAAGAC 1562  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 1563 AATGCCCACTCTGTGACGACCCCTTCCAGGCTACTGTCTGTGAGAGAGCTCCCTTA 1622

QY	521	GLYTYRLeuValIleuHISValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu	540	QY	881	ArgGluAsnValAlaGlnTyRValIleuArgAlaTyRAlaValAspLysGlyMetPro	900
Db	1623	GGCTACTGCTGTTCTCATGTCCAGGCTATCGACGATGCTGTCAGCAATGCCCTG	1682	Db	2703	CGAGAGAAAGTGGCCAGATATCTTGGGGCATATGACATGAGCAAAAGGAGATGCCCA	2762
QY	541	GLUTYRArgLeuAlaGlyValGlyHISAspPheProPheThrLISAsnGlyLysArgLys	560	QY	901	AlaArgThrProMetGluValThrValThrValIleuAspValAsnAspAsnProProVal	920
Db	1683	GAATACCGCTGCTGGGGTGGAGCAATGACTTCCCTTCCATCAACATGAGCAAGGC	1742	Db	2763	GCCCGCACACTATGAGAAAGTACAGTCACTGTGGTGGATGTGAATGACAAATCCCCCTGTC	2822
QY	561	TPRLeuSerValAlaAlaGluLeuAspArgGluGluValAspPheTyRSerPheGlyVal	580	QY	921	PheGluGlnAspGluPheAspValPheValGluGluAsnSerProLISGlyLeuAlaVal	940
Db	1743	TGGATCTGTGGCTGCTGACCTGACCGGAGAGAGTTGATTCTACAGCTTTGGGGTA	1802	Db	2823	TTTGAAGCAGATGACTTTGATGTGTGTGGAGAGAAACGCCCATTTGGGCTTACCGCTG	2882
QY	581	GluAlaArgAspHISGlyThrProAlaLeuThrAlaSerAlaSerAlaSerValThrVal	600	QY	941	AlaArgValThrAlaThrAspProAspGlyLysIleThrAsnAlaGlnIleMetTyRLeuIle	960
Db	1803	GAAGCTCGAGACCATGTGCACTCCAGACACTGCTCGGCTCGAGTCCAGCGTGACTGTGC	1862	Db	2883	GCCCGGGTACACAGCACTGACCCCGATGAGGACCAATGCCCATATATATGACAGATT	2942
QY	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyRThrValArgLeuAsn	620	QY	961	ValGluGlyAsnIleProGluValPheGluLeuAspLISepheSerGlyLysLeuThrAla	980
Db	1863	CTGGATGTCACACACAAATCCATCTTACCCAAACAGATGACAGTGGCGGCTCAAT	1922	Db	2943	GTTGAGGGGACATCCCTGAGGTCCTTCCAGCTGACATCTTCCGGGGAGCTGACAGCC	3002
QY	621	GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHIS	640	QY	981	LeuValAspLeuAspTyRValAspArgProGluTyRValIleuValIleGlnAlaThrSer	1000
Db	1923	GAGGATGCACTGTGGGCAACAGCGTGGTCAAGCTGTGACCGTGAATGCTCAT	1982	Db	3003	CTGGTACCTTGAATTAAGAGGACCGGCTGATACGTCTGTGATCCAGGCCACGTCA	3062
QY	641	SerValIleThrTyRGlulIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660	QY	1001	AlaProLeuValSerArgAlaThrValHISValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	1983	AGTGCATCACTTACCAATCCAGATCCAGTGGCAATATCTGAAACCGCTTCTCATCAACAC	2042	Db	3063	GCTCCTCTGTGAGCCGGGCTTACAGTCCAGCTCCGCTCTTGAACCGCAATGACAAACCA	3122
QY	661	GlnSerGlyGlyLysLeuValSerLeuAlaLeuProLeuAspTyRLeuGlnLysArgLys	680	QY	1021	ProValIleuGlyAsnPheGluIleuPheAsnAsnTyRValThrAsnArgSerSer	1040
Db	2043	CAAAAGTGGTGGGCTGGATTCCTTGGCTGGCCACTGACATCAAACTTGAAGCGGCGAG	2102	Db	3123	CGATGCTGGGACATTTGAGATCTTTCACAACTATGTCCAAATGCTCAAGCTCAAGCGC	3182
QY	681	TyRValIleuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700	QY	1041	PheProGlyGlyAlaIleGlyArgValProAlaHISAspProAspLISerAspSerLeu	1060
Db	2103	TATGTGTTGGCTGTATCCGCTCCGATGGCACTCGGCAAGACACGGCAAGATTGGGTG	2162	Db	3183	TTCCCTGGGGGTGCCATTGGCGAGATCACTGCCATGACCTGATATCTCAGATGCTGTG	3242
QY	701	AsnValThrAspAlaAsnThrHISArgProValPheGlnSerSerHISTyRThrValAsn	720	QY	1061	ThrTyRSerPheGluArgGlyValAsnGluLeuSerLeuValIleuLeuAsnAlaSerThrGly	1080
Db	2163	AATGTCAACGAGCGCCAAACCCATCTCTCTTCCAGAGCTCCCACTATACAGGAAAT	2222	Db	3243	ACTTACACCTTTGAGCGGGGAAATGAACTCAGCTGGTCTCGTCCAAATGCTCCACGGGT	3302
QY	721	ValAsnGluAspArgProAlaGlyThrThrValIleuIleSerAlaThrAspGluAsp	740	QY	1081	GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	2223	GTTAATGAGAACCGGCGGCGGCAAGCACAGGTGGTCTATGACGCCCAAGATGAGGAC	2282	Db	3303	GAGCTGAGCTTAAAGCCGGGCACTGGACAAACACGGCTCTGGAGGCACTATGAGCGTG	3362
QY	741	ThrGlyGluAsnAlaArgIleThrTyRPhMetGluAspSerIleProGlnPheArgLys	760	QY	1101	LeuValSerAspGlyValHISerValThrAlaGlnCysAlaLeuArgValThrIleIle	1120
Db	2283	ACAGGTGAGAAATGCCCGATCACTTACTCATGAGAGACAGCATCCCAAGTTCGCGATC	2342	Db	3363	CTGGTGCAGACGGGTGATCACAGGTGACCGGCCAGGTGCGGCTGCGTGAATCATATC	3422
QY	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyRGlulAspGlnValSer	780	QY	1121	ThrAspGluMetLeuThrHISerIleThrLeuArgLeuGluAspMetSerProGluArg	1140
Db	2343	GATGACAGACCGGGGCTGTCAACCCAGGCTGAGCTGACATCCAAACCAAGAGTCT	2402	Db	3423	ACCGATGAGATGCTCACCCACAGATCACTGCGGCTGGAGAGCATGTCAACCCGAGCGC	3482
QY	781	TyRThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrTyR	800	QY	1141	PheLeuSerProLeuLeuGlyLysPheIleGlnAlaValAlaIleThrLeuAlaThrPro	1160
Db	2403	TACACCTTGGCCATTAATCTCTGGGACAAATGGATTTCCCAAGAGTCCGACACACTAC	2462	Db	3483	TTCTCTGATCACTGTGATGAGCTCTTCAATCAAGGGGTGGCGGCACTGACAGCGCA	3542
QY	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyR	820	QY	1161	ProAspHISValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHISIle	1180
Db	2463	CTGAGAGATCTGTGTGACACAGCGTGAATGACATGCCCTCAAGTTCCTGGAGAACTCTAC	2522	Db	3543	CCGAGCACAGTGGTGTCTTCAAGTACAGGGGACACCGACGCCCGGGGGGCCACATC	3602
QY	821	GlnGlySerValTyRGlulAspValProProPheThrSerValIleuGlnIleSerAlaThr	840	Db	3603	CTCAACGTGACCTGTGTGTGGTGGCCAGCGCCAGGGGCCCGGGGGGCCCTTCTGTG	3662
Db	2523	CAGGGGACGTCTATGAGATGTGCCACCTTCACTAGAGGTCTCTGAGATCTCACGCACT	2582	QY	1201	ProSerGluAspLeuGlnGluArgLeuTyRLeuAsnArgSerLeuLeuThrAlaIleSer	1220
QY	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyRThrPheGlnGlyGlyAspAspGly	860	Db	3663	CCCTTGAAGACCTGACAGAGCGCTTATCTTCAACCGCACTGTGCTGACGGCATCTCG	3722
Db	2583	GATCTGATCTTGGACTTAAATGGCAGAGTCTTCTACACTTCCAAAGAGGCAACATGGA	2642	QY	1221	AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGlyAsnTyR	1240
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp	880	Db	3723	GCAAGCGCGTGTGCCCTTGCATGACAGACAACTGTGCGGGAGGCCCTGGAGAGACTAC	3782
Db	2643	GACGGTGACTTATATGTGATGATCCAGTCAAGGATGTGTGAAACGCTTACGAGGGCTGAT	2702	QY	1241	MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer	1260

[illegible]

Db	4863	CCCTGG3CTTTGGGGGCAAGAGCTGGGCCCAAGAAATGGCAATCCAGACACTTCTG	4922
QY	1621	GIYserSerIeuVal1aATPHisgLYIleusSerIeuProIIeserGInProTPHYIleu	1640
Db	4923	GGCAGAGCCCTGGTGGCTGGCAAGGGCTCTGGCTGCCATCTCCAACTGTACTTC	4982
QY	1641	SerIeuNeuPheATgThrATgGln1aAspG1yValIeuIeuGln1a1IethATgG1y	1660
Db	4993	AGCTCATGTTTCCGACCGCCGACGGCCGACCGGTCTCTGCTGACGGCCATTCACAGGGGG	5042
QY	1661	ArgSerThrIIethIleuGlnIleuATgGlnIyHisValIeIleuSerValGluG1YThr	1680
Db	5043	CGCAGACCCATCACCTTCACAGCTACGAGAGGGCCACGTATGCTTGACCTGGAGGGCA	5102
QY	1681	G1yIleuGln1aIaserSerIeuATgIleuGlnProG1yATg1aAanaPGLYAspTTHis	1700
Db	5103	GGGCTTCAAGCTCTCTCTTCCTGGCTGGAGCCAGGCCCGGCCATTCAGGTGATCGGAC	5162
QY	1701	His1aGlnIeu1aIeuG1y1aIaserG1yG1yProG1yHis1a1IeIleuSerPheAsp	1720
Db	5163	CATGCACAGCTGGCACTGGAGGACAGGGGGGGCTGGCCATGTCATTCTTCCTGCAT	5222
QY	1721	TYTg1yGInGlnATg1aGluG1yAsnIleuG1yProATgIeuHisgLYIleuHisIleuSer	1740
Db	5223	TATGGGCGAGAGAGAGAGAGGGGCACTGGGCCCCGGCGCAATGCTTCACACTGAC	5282
QY	1741	Asn1IethVal1aG1yG1y1IeProG1yPro1aG1yG1yVal1a1aATgG1yPheATgG1y	1760
Db	5283	AACATACAGTGGGCGGAAATCTCGGGCCAGCCGGCGGTGTGGCCGTTCGGGGG	5342
QY	1761	CysIleuGlnG1y1a1aATgValSerAspThrProGluG1yValaAsnSerIeuAspProSer	1780
Db	5343	TGTTTGGCAGGGTGTGGCGGGTGAAGGATACGCAAGGGGGGTTTACAGCTGGAAATCCAGC	5402
QY	1781	Hisg1yG1uSer1IeanaVal1aGluG1yG1yCysSerIeuProAspProCysAspSerAsn	1800
Db	5403	CATGGGAGAGACATCAGACGTGGAGCAAGGCTGTAGCTGCTGACCTGTGATCTCAAC	5462
QY	1801	ProCysPro1aAsnSerTYrCysSerAsnAspTTPAspSerTYrSerCysSerCysAsp	1820
Db	5463	CCGTGCTCGCTTACAGCTATTGGAGCAACGACTGGACAGCTTTCCTGACGCTGTGAT	5522
QY	1821	ProG1yTYrTYrG1y1aAspAsnCysThrAsnValaCysAspIeuAsnProCysGlnHisGln	1840
Db	5523	CCAGGTATCTATGGTGAACACTGTACTATATGTGTGACTGAACCGGTGTGAACAG	5582
QY	1841	SerValCysThrATgIyPProSer1aProHisgLYrTYrThrCysGluCysProProAsn	1860
Db	5583	TCTGTGTATACCCGCAAGCCCAAGGCCCCCATGGCTATACCTCGAGATGCCCAAT	5642
QY	1861	TYrIleuG1yProTYrCysGluThrATg1IeAspGlnProCysPro1aG1yTYrTTPG1y	1880
Db	5643	TACCTTGGGCAATCTGTGAGACACAGATTCACAGCCCTGTCCCGGTGTGTGGGA	5702
QY	1881	HisProTHrCysG1yProCysAsnCysAspValSerIyG1yPheAspProAspCysAsn	1900
Db	5703	CATCCCAATGTGGCCCAATGCATGTGATGTACGAAAGGCTTGTGAACCAAGATCGAAC	5762
QY	1901	LYrThrSerG1yGluCysHisCysIyG1uAsnHisTYrATgProProG1ySerProThr	1920
Db	5763	AAGCAAGACGGCGAGTGCACCTGCAGAGAGAACATACCGGCCCCAGGAGGCCAC	5822
QY	1921	CysIeuIeuCysAspAspTYrTYrProTHrG1ySerIeuSerATgVal1yAspProGluAsp	1940
Db	5823	TGCTCTTGTGTGACTGTACCCCAAGGCTCTTGTCCAGATCTGTGACCTGAGAT	5882
QY	1941	G1yGlnCysProCysIyAspProG1yVal1IeG1yATgGlnCysAspATgCysAspAsnPro	1960
Db	5883	GAGCAGTGTCAATCAAGCAGGTGTATCATGGGGTCAAGTGTGACCGCTGTGAACAACCT	5942
QY	1961	Phe1aG1yVal1aTHrThrAsnG1yCysGluValaAsnTYrAspSerCysProATg1a1Ie	1980
Db	5943	TTGTGTAGGTACACCAATGGCTGTGAATGTAAATTTATGACAGGTGCCACAGAGCAT	6002

QY	2341	ValValAlPheaArgAmGInuSerHisValSerCyseGInuCyAsanHisSmetThrSerPheAla	2360
Db	7083	GTGCTCTCCGCAATGAGAGCCAGCTCAAGCTGCCAGTGCACCAACATGACAGAGCTTGCTG	7142
QY	2361	ValIleuMetAspValSerArgArgGluAsnGlyLuiIleuProLeuYsThrLeuThr	2380
Db	7143	GTGCTCAATGAGAGTTTCTCGGCGGGAGAGATGGGGAGATTCCTGCACATGAAACACTGACA	7202
QY	2381	TyrValAlaLeuGlyValAlaThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu	2400
Db	7203	TACGGGTGCTCAGAGGTGCACCTTGTGGCCCTTGTGTCAACCTTCTCTTCTCTCACTCTC	7262
QY	2401	LeuArgIleLeuAspSerSerAmGlnHisGlyIleArgArgAmLeuThrAlaAlaLeuGly	2420
Db	7263	TTGGGTATTCCTGGGCTCCAAACCAACAGCGATCCGATCCGATCAAGACGTGGCTGGGG	7322
QY	2421	LeuAlaGInLeuValPheLeuLeuGlyYIleAsnGlnAlaAspLeuProPheAlaCysThr	2440
Db	7323	CTGGCTCAGCTGAGTCTTCTCTCTGGGAGATCAACAGAGCTGACCTCCCTTTTGGCTGACA	7382
QY	2441	ValIleAlaIleLeuLeuHisPheLeuYrLeuCyThrPheSerTrpAlaAlaLeuGlu	2460
Db	7383	GTCAATTGCCATTCCTGCTGCACTTCTGTACTCTGCAACCTTTCCTGAGGCTGCTGAG	7442
QY	2461	AlaLeuHisIleuYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Db	7443	GCCTTGACCTGTACCGGGGACCTCACTAGAGGTGGCATGTGCAACACCGGGCCCATCGC	7502
QY	2481	PheYrYrMetLeuGlyYrTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Db	7503	TTCTACTACATGCTGGGCTGGGGGTGGCTGCTTATCAACAGGCTACAGCGTGGGGCTG	7562
QY	2501	AspProGluGlyYrGlyYAsnProAspPheCysTrpLeuSerIleYrAspThrLeuIle	2520
Db	7563	GACCCCGAGGGCTACGGGAAACCTGCACTTCTGCTGGCTCTCCATCTATGACAGCTATAC	7622
QY	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuYrIleu	2540
Db	7623	TGGAGTTTGTGTGCCCGGAGGGCTTTGGCCGTCCGATGAGTGTCTTCCGTACATCTG	7682
QY	2541	AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlyYsIleYsGlyProValSer	2560
Db	7683	GGGGCCGGGGCTCTCTGTGCTGGCCAGCGGAGGCTTTGAGAGAAAGGTCTGTCTTG	7742
QY	2561	GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu	2580
Db	7743	GGCTGTGAGGCTCTCTTCCGCCGTCTCTCTGCTGTGAGGCCACAGTGGCTGTGACCTG	7802
QY	2581	LeuSerValIleSerAspThrIleuLeuPheHisIleYrLeuPheAlaThrCysAsnCysIle	2600
Db	7803	CTCTCTGTCAACAGCGACACCTCTCTCTTCCACTCACTCTTGTGTACCTGCATATGATC	7862
QY	2601	GlnGlyProPheIlePheLeuSerYrValValLeuSerYsGlnValArgGlyAlaLeu	2620
Db	7863	CAGGGCCCTTCAATCTTCTCTCTCTATGTGTGTGTATGACAGAGAGGTCCGGAAAGCATC	7922
QY	2621	LysLeuAlaCysSerArgGlyAspProSerProAspProAlaLeuThrTrpYsSerThrLeu	2640
Db	7923	AAGCTTCCCTGACGGCCGCAAGCCCAAGCCCTGTGACCTCTGTGACCAACCAAGTCCACCTG	7982
QY	2641	ThrSerSerYrAsnCyProSerProYrAlaAspGlyArgLeuYrGlnProYrGly	2660
Db	7983	ACCTCGTCTTACACTGCCCAAGCCCTTACGCAAGATGGGGGCTGTACCAAGCCCTTACGA	8042
QY	2661	AspSerAlaGlySerLeuHisIleSerThrSerArgSerGlyLysSerGlnProSerYrIle	2680
Db	8043	GACTCGGCGGCTCTCTGTGACACACCACTGACGTCTCGGGCAAGAGTCAAGCCCACTACATC	8102
QY	2681	ProPheLeuLeuArgGluGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly	2700
Db	8103	CCCTCTTGTGAGAGGAGAGTCCGCACTGAACTTGGCCCAAGGGCCCTCTGTGGCTGGGG	8162
QY	2701	AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer	2720

DB: 8163 GATCCAGGAGGCTGTTCCCTGGAGAGCTCAAGACAGCATATATCTGACACGAGCTCC 8222  
QY 2721 AASPSeAAspLeuSerLeuGluAAspAAspGInserGlySerThrAlaSerThrHisSerSer 2740  
DB 8223 GACAGTACCTGCTCTTAAAGACACAGAGTGGCTCTTAATGCTCTTAACTCATCA 8282  
QY 2741 AASPSeGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
DB 8283 GACAGTACGAG 8342  
QY 2761 TTPAAspSerLeuLeuGlyProGlyAlaGluAAspProLeuHisSerThrProGlyAAsp 2780  
DB 8343 TGGAGTACGCTGCTGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8402  
QY 2781 G1YGLYProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 2800  
DB 8403 GGGGGCCCAAGGCTTGGAG 8462  
QY 2801 SerSerGlyAAspGlyAAspProGluAAspGluAAspGluAAspGluAAspGluAAsp 2820  
DB 8463 AGTAAAGTGGACAG 8522  
QY 2821 G1UGLYSerLeuGlyProLeuProGlySerSerAlaGlnProHisGlyGlyLeuLeu 2840  
DB 8523 GAGGGGTCCCTAGGAG 8582  
QY 2841 LysLysCysLeuProThrHisSerGlyLysSerSerLeuAAspProLeuGluGlu 2860  
DB 8583 MAGAAGTCTGCTCCCAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642  
QY 2861 CysThrGlySerSerAlaGlySerSerAlaSerGlyLysSerGlyProProPro 2880  
DB 8643 TGGACAGAGGCTTCCCGGGGCTCTCCGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 8702  
QY 2881 ArgProProProArgGlnSerLeuGlnGluGluGluGluGluGluGluGluGluGlu 2900  
DB 8703 CGCCCAAGGCTTCCCGGGGCTCTCCGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 8762  
QY 2901 SerLysLysAlaGlyThrValAAspGluAAspSerSerGlySerGluPheLeuPheAAsp 2920  
DB 8763 AGCATCAAG 8822  
QY 2921 PheLeuHis 2923  
DB 8823 TTCCTGCAAT 8831

RESULT 3  
US-09-788-711A-1  
; Sequence 1, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Teesta  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788, 711A  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 8871  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-1

Alignment Scores:  
Pred. No.: 0  
Score: 2837.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 97.06%

Length: 8871  
Matches: 2837  
Conservative: 0  
Mismatch: 0  
Indels: 0

DB: 9 Gaps: 0  
US-09-916-849a-3 (1-2923) x US-09-788-711A-1 (1-8871)  
QY 1 MetAAspSerProAlaThrGlyValProLeuProProProProProProLeuLeuLeu 20  
DB 1 ATGCGAGAGCCCGGACACCGGCGTCCCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAAspGlnValGlyProCysAAspSerLeu 40  
DB 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
QY 41 GlySerAAspGlyAAspGlySerSerGlyAAspAAspAAspAAspAAspAAspAAspAAsp 60  
DB 121 GGGTCAAG 180  
QY 61 SerAlaSerAAspLeuThrLeuThrThrSerArgCysAAspAAspAAspAAspAAspAAsp 80  
DB 181 TCAAGGTGGAACCTTGCTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 81 GlyHisLeuValProHisHisAAspGlyLeuAAspValTTPCysProGlySerGluAAsp 100  
DB 241 GGCACACTGTAACCCACACAGATGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
QY 101 TLeuProLeuProProAlaProGlyLysCysProTTPSerCysAAspLeuLeuGlyTLeu 120  
DB 301 ATTCCCTTACACACAGCTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 121 GlyHisLeuSerProGlyLysLeuThrLeuProGlyLysLeuValProCysLeuValAAsp 140  
DB 361 GGCACACTTCCCAAG 420  
QY 141 ProAAspLeuAAspCysGlnSerCysLysLeuAAspGlnAAspGlyLeuAAspGlyLys 160  
DB 421 CCAAGGCTCAAGTGCAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 161 ArgSerProGlyLysLeuGlyLysLeuGlyLysLeuGlyLysLeuGlyLysLeuGlyLys 180  
DB 481 AGGTCAACCAAG 540  
QY 181 PheGlnProProSerThrGlnAlaThrValProGluAAspGlnProAlaGlyThrProVal 200  
DB 541 TTCAG 600  
QY 201 AlaSerLeuAAspAlaAAspProAAspGlyLysAAspGlyLysAAspGlyLysAAspGly 220  
DB 601 GCATCCCTGAG 660  
QY 221 AAspAlaLeuPheAAspSerArgSerAAspGlnPhePheSerLeuAAspProValThrGly 240  
DB 661 GATGCCCTCTTGAATGAGCGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 241 ValThrThrAlaGluGluLeuAAspArgGluThrLysSerThrHisValPheArgValThr 260  
DB 721 GTAACCAAG 780  
QY 261 AlaGlnAAspHisGlyMetProArgAAspSerAlaLeuAlaThrLeuThrLysLeuValThr 280  
DB 781 GCGAG 840  
QY 281 AAspThrAAspHisAAspProValPheGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
DB 841 GACACCAATGACCAATGACCTGTGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 301 AAspLeuGluValGlyTyrGlyValLeuThrValAAspAlaThrAAspGlyAAspAlaProPro 320  
DB 901 AACCTGAGAGTGGCTTAAAG 960  
QY 321 AAspAlaAAspLysLeuThrArgLeuLeuGlySerGlyLysProSerGlyValPhe 340  
DB 961 AATGCAATATTCGTACCGCTGAG 1020  
QY 341 GlnLysAAspProArgSerGlyValLysLeuThrArgGlyProValAAspArgGlyLysVal 360

```

Db      1021 GAGATGACCCCTCGCTCTGGGTGATCCGAAACCGGCGCTGTGATCGGAGAGAGTG 1080
Qy      361 GJuseTyrGlnLeuThrValGluIleAspArgGlnGlyArgAspProGlyProAspSer 380
Db      1081 GAATCTTACAGCTGACGGTAGAGGACAGAGTACAGAGTCCGAGTCCGAGT 1140
Qy      381 ThrThrAlaValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPheSer 400
Db      1141 ACCAGCGCGCTGTTTCTTCTTCTGTGGAGATGACATGATATCCCTCCAGTTAGT 1200
Qy      401 GJuvAspGlyTyrValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
Db      1201 GGAAGCGGTATGTGTCCAGGTGAGGAGATGATCTCCAGGGCCCCAGTACTCCAA 1260
Qy      421 ValThrAlaSerAspArgAspGlySerAspAlaValIleIleTyrSerIleMetSer 440
Db      1261 GTCAACGCTCGGATCGAGACAGGGAGCATGCGGTGTGCACTATAGCATCATAGAGT 1320
Qy      441 GJyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
Db      1321 GGCAATGCTCGGGGACAGTTTATCTGATGCCCAAGCTGGAGCTTGATGTGGTGAAC 1380
Qy      461 ProLeuAspTyrGlnThrThrLysGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db      1381 CCTCTTGACTATGAGACGACCAAGAGATACCCCTACGGGTGCGACACAGATGTGTGC 1440
Qy      481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db      1441 CTTCCCTCCCTCTCTATATGCTGTGGCTGTGTGACATGAGTCTGTGATTCAAACGAC 1500
Qy      501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      1501 AATGCCCATCTTGTGTGTCAGACCCCTTCCAGGCTACTGTCTGTGAGAGCGTCCCTTA 1560
Qy      521 GJyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db      1561 GGCTACCTGGTCTTCATGTCCAGGCTATCGACGCTGATGCTGTGACATGCGCCCTG 1620
Qy      541 GJyTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1621 GAATACCGCTTGTGTGGGTGGACATGACTTCCCTTCCATCAATCAATGACACAGGC 1680
Qy      561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      1681 TGGATCTCTGTGGCTGTAACCTGACCGGACCGGAGAAAGTTGATTTTCAAGCTTTGGGATA 1740
Qy      581 GJuaIleArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThrVal 600
Db      1741 GAAGCTCGAGACATGACCTCCAGCACTCACTGCTCGGCAAGTGTCAAGCTGACTGTC 1800
Qy      601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      1801 CTGGATGTCAAGACACACATCAACCTTTACCAACCAAGTACACAGTGGCGGCTCAAT 1860
Qy      621 GJuaAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      1861 GAGGATGACGCTGTGGGACCAAGCGTGTGACGGTGTCAAGTGTGAACCGTATGTCTAT 1920
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1921 AATGTCATCACTTACCAATCAACAGTGCATATCCGAAACCGCTTCCATCAACAGC 1980
Qy      661 GlnSerGlyGlyValLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln 680
Db      1981 CAAAGGTGGTGGGTGTATCCCTTGTGCTGCCCTGCACTGCACTCAAACTTAAGGGGACG 2040
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      2041 TATGTGTGGCTGTATACCGCTCCGATGACACTCGGACAGACACGGCACAATGTGGTG 2100
Qy      701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720

Db      2101 AATGTCACCGACCAACACCCATGCTCTTCTTCAAGAGTCCCACTATACATGAAAT 2160
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValIleIleSerAlaThrAspGluAsp 740
Db      2161 GTTAATAGAGACCGGCGGACAGACACACGGTGTGTATACAGCCACAGATGAGGAC 2220
Qy      741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2221 ACAGGTGAAGATGCCCGATCACTTCACTTCAAGAGACAGATCCCCAGTTCCGCATC 2280
Qy      761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2281 GATGCAGACACGGGGCTGTCAACCAAGGCTGAGTGCATACAGAACCAAGTGTCT 2340
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db      2341 TACACCTTGGCCATTAAGCTCTGGGACATGCAATTCCTCCAAAGTCCGACACACTAC 2400
Qy      801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db      2401 CTGAGATCTCTGTGACAGCGTGAATGACATATGCCCTCAGTTCTGTGAGACTCTTAC 2460
Qy      821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2461 CAGGCAAGTGTCTATGAGAGATGTGCACCTTCACTACGCTCTGCAAGTCTCAAGCT 2520
Qy      841 AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyValAspAspGly 860
Db      2521 GATCGTATCTTGACCTTAATGACAGGCTTCTTACACTTCCAAAGAGGCCACATGGA 2580
Qy      861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db      2581 GACGCTGACTTATATGTGATGCCACGTCAAGGCATCGTGCAGACGCTACGAGGCTGAT 2640
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyGlyMetProPro 900
Db      2641 CGAGGAACGTGGCCCAATATGTCTTGGGGATGTGACAGTGAACAGGGAGATGCCCA 2700
Qy      901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db      2701 GCCCGCACTTATGGAAGTACATCACTGTGTGATGTGAATGACATATCCCTGTTC 2760
Qy      921 PheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2761 TTGAGCAGATGAGTTGATGTGTGTGAAAGAGACAGGCCATTTGGGCTACCGCTG 2820
Qy      941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      2821 GCCCGGTCAAGACCACTGACCCCGATGAGGACCAATGCCCAGATTATGTACCAAGAT 2880
Qy      961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db      2881 GTGAGGGCAACATCCCTGAGGCTTCCAGCTGACATCTTCTCGGGAGCTGACAGCC 2940
Qy      981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db      2941 CTGGTACACTTACAGTACAGGACCGGCTGATGACGTCCTGTATCAAGGCCACGCA 3000
Qy      1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspAsnPro 1020
Db      3001 GCTCCTGTGTGAGCGGGCTTACAGTCCAGTCCGCTTGAACCGCAATGACAAACCA 3060
Qy      1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db      3061 CCAAGTGTGGGCACTTTGATATCTTTCAACAACATATGTCAACAATCGCTCAAGACGC 3120
Qy      1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3121 TTCCCTGGGGGTGCATTGGCCGAGTACTGCCATGACCTCGATATCTCAGATGATGTG 3180
Qy      1061 ThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      3181 ACTTACAGCTTGAAGCGGGAAATGAATCAAGCTGTGTCTCAATGCTTCCACGGGT 3240

```



QY 1081 GIuleuLYeLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
DB 3241 GAGCTGAAGCTTAAGCCGGCGCACTGGACAACAACCGGCTCTTGGAGGCCATCATGAGCGTG 3300  
QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
DB 3301 CTGGTCTGAGACGGCTGTACACAGCTGTACCGCCAGTGGCGCTGTGTGTGACATCATC 3360  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
DB 3361 ACCGATGAGATGCTTCCACCAACACATACGCTGCGCTGGAGGACATGTCAACCGAGCGC 3420  
QY 1141 PheLeuSerProLeuLeuGlnLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
DB 3421 TTCCTGTCAACACTGCTAAGGCTCTTCACTCAGCGCGTGGCCGCAAGCTGGCAAGCCA 3480  
QY 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyValIleIle 1180  
DB 3481 CCGGACCAAGTGGTGTTCACACGTACAGCGGGAACAACGCCCGCGGGGCCACATC 3540  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyValGlyProPheLeu 1200  
DB 3541 CTCAAGTGAAGCTGTGCGGTGGCCAGCGCCAGGGCCCGGGGCCGCGCTTCTCTG 3600  
QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
DB 3601 CCTCTGAGGACTGGCAGAGCGCTATACCTCAACCGAGCCTGTGACGCGCATCTCG 3660  
QY 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240  
DB 3661 GGCACACGGGTGTGCTTCCCTTGCAGACCAACATGTGCTGGGGAGCCCTGCGAGACTAC 3720  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
DB 3721 ATCCGCTGCGTGTGCGGTGCTGCGCTTGCATCTTCCGCGCTTCAATCGCTCTCTCC 3780  
QY 1261 ValLeuPheArgProIleHisProValGlyValLeuArgCysArgCysProProGlyPhe 1280  
DB 3781 GTGCTCTTCCGGCCCAATCCACCCGTCGAGGGCTCGCTGCGCTGCGCGCGCTTCC 3840  
QY 1281 ThrGluAspTyrCysGluThrGlnValAlaPheLeuCysTyrSerArgProCysGlyProHis 1300  
DB 3841 ACCGAGTACTACTGCGAGACCGAGGTGTGACTGTGCTGCGCGCTGTGGCCCCCAC 3900  
QY 1301 GlyArgCysArgSerArgGlnGlyTyrTyrCysLeuCysArgAspGlyTyrThrGly 1320  
DB 3901 GGGCGTGGCGGAGCGCGCGAGGGCGGCTACACTGCTCTGTGTGTGTGTGTGTGTGTGT 3960  
QY 1321 GlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGly 1340  
DB 3961 GAGCAGTGTGAGGTGAGTGTCTCGCTCAGCGCGCTTGCACCCGGGTGTGTGTGTGTGTGT 4020  
QY 1341 GluThrCysValAlaAsnLeuValGlyValPheValCysAspCysProSerGlyAspPhe 1360  
DB 4021 GGGACCTGTGTCAACCTGT 4080  
QY 1361 GluValProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
DB 4081 GAGAAAGCCCTACTGCGAGGT 4140  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGluArg 1400  
DB 4141 CCGGGCTGGCGAGCGCTTTCACCTTCACTGCGCTTCTGTGTGTGTGTGTGTGTGTGTGT 4200  
QY 1401 AspGlyLeuLeuLeuTyrArgGlnArgPheAsnGluValHisAspPheValAlaLeuGlu 1420  
DB 4201 GAGCGGT 4260  
QY 1421 ValIleGlnGlnValGlnLeuThrPheSerAlaGlyValLeuSerThrThrValSer 1440  
DB 4261 GTGATCCAGAGGAGGT 4320

QY 1441 ProPheValProGlyValValSerAspGlyGlnThrHisThrValGlnLeuLeuTyrTyr 1460  
DB 4321 CCATTCGTGGCCGGAGAGGTCAAGTATGTGCGCAGGTGTGTGTGTGTGTGTGTGTGTGTGT 4380  
QY 1461 AsnIleProLeuLeuGlnGlnThrGlyLeuProGlnGlyProSerGlnGlnIleValAla 1480  
DB 4381 AATTAAGCCACTGT 4440  
QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
DB 4441 GTGTGTACCGT 4500  
QY 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlySerIleValSerLeuAspLeuThr 1520  
DB 4501 AACTACTCTGT 4560  
QY 1521 GlyProLeuLeuLeuGlnGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
DB 4561 GGGCCCTGT 4620  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
DB 4621 CAGTTCGTGGCTGT 4680  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaValValAsnValCysAspSer 1580  
DB 4681 TTCATTGTCCAAATGAGCAGGT 4740  
QY 1581 AsnThrCysHisAsnGlyValThrCysValAsnGlnThrPheProAlaPheSerCysGluCys 1600  
DB 4741 AACACTTGTCCAAATGAGGAGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
QY 1601 ProLeuGlyPheGlyGlyValSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
DB 4801 CCCCTGGGCTTTGGGGGGAAGAGCTGGCGCCAGAGAAATGGCAATCCACAGACATCTTCTG 4860  
QY 1621 GlySerSerLeuValAlaIleThrHisGlyLeuSerLeuProIleSerGlnProTyrIleu 1640  
DB 4861 GGCACACAGCTGT 4920  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
DB 4921 AOCCTCATGTTCCGACCGCGCAGGCCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGluValThr 1680  
DB 4981 CGCACACACATCACCTTACAGTACGAGAGGGCCAGGATGATCTGACGCTGTGAGGGCACA 5040  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAlaAsnAspGlyAspThrHis 1700  
DB 5041 GGGCTTCAAGGCTCTCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100  
QY 1701 HisAlaGlnLeuAlaLeuGlnValAlaSerGlyProGlyHisAlaIleLeuSerPheAsp 1720  
DB 5101 CATGTACAGCTGT 5160  
QY 1721 TyrGlyGlnGlnArgAlaGlnGlyValAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
DB 5161 TATGGGACAGAGAGAGAGAGGCAACTGTGGCCCCCGGCTGTGTGTGTGTGTGTGTGTGT 5220  
QY 1741 AsnIleThrValGlyValIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly 1760  
DB 5221 AACATTAACAGT 5280  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAlaSerSerLeuAspProSer 1780  
DB 5281 TGTGTGTGAGGAGGT 5340  
QY 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
DB 5341 CATGGGAGAGCATCAAGCTGT 5400  
QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp 1820



Db	5401	CCGTGCTCTGCTAAGCTATTGACAGCAAGCTGGAGACAGCTATTCTGCAAGCTGTGAT	Db	6481	ACCATGTCAGCGCCCAACATTGTCACTCCGTAGTGCCTTGGAACAAAGGAACCTTGCT	6540
Qy	1821	ProglYTYrTgLYaApAnCySThrAaenValCYaAPLeuAaenProCYaSLuHISgln	Qy	2181	GLyAlaLYLeuProArgTYrGLuAlaLeuArgLYglnGlnProProAaenLeuHlurh	2200
Db	5461	CCAGGTACTATGAGTGAACAAGCTGTAATATGTGTGACTGAACCCGTGTAGACACAG	Db	6541	GGGGCCAAAGCTGCCCGCTGACAGGCTTGCGTGGGAGAGACCCCGGACCTTAGACA	6600
Qy	1841	SerValCYeThThArgLYaPProSerAlaProHISgLYTYrThCYaSLuCYaProProAaen	Qy	2201	ThrValILEuProGlnSerValPheArgLYurhProProValValArgProAlaGly	2220
Db	5521	TCGTGTGTATCCCGCAAGCCCAAGTGCCTCCCATGTGCTATACCGAGAGTGTCCCCCAAT	Db	6601	ACAGTCATTGCTGAGTGTCTTTCAGAGAGAGCCCGCGTGTGAGGCCCGGAC	6660
Qy	1861	TYrLeuGLYProTYrCYaSLuHlurhArgLYLeaProGlnProCYaProArgLYrTYrTYrGLY	Qy	2221	ProGLYglnAlaGlnGlnProGLYglnLeuAlaArgLYglnArgLYHISglnLeu	2240
Db	5581	TACCTTGGGCAATCTGTAGACACAGGATGTACAGACCTGTGCTCCGTGGCTGGGGGA	Db	6661	CCCGAGAGGCGCCAGAGACCAAGAGGCTGGCAAGCCGACACAGCCAGCCGAGAGCTG	6720
Qy	1881	HisProThrCYaSLYProCYaAaenCYaAaenValSerLYgLYPheAaenProAaenCYaAaen	Qy	2241	SerGlnLYglnAlaValAlaSerValILETYrArgTYrThLeuAlaGLYLeuAaenPro	2260
Db	5641	CATCCCAAGTGGGCGCAAGCTGTATGTACAGAAAGGCTTTGACCAACCTGACACCTGCAAC	Db	6721	AGCCAGGCTGAGGCTGTGGCCAGGCTTCATCTACCCGACCTTGCCGGGCTATTGCTT	6780
Qy	1901	LYeThrSerGLYglnCYaHISCYaLYglnAaenHISTYrArgProProGLYSerProThr	Qy	2261	HisAenTYrAaenProAaenPheArgLYaSerLeuArgLYaProLYaArgProILEaAenThr	2280
Db	5701	AAGACAAAGCGGAGTGCACCTGCAAGAGAAACCACTACCGGCGCCCAAGGCAAGCCCAAC	Db	6781	CATAACTATGACCTTGACAGAGCGAGCTTGAGAGTCCCAAAACGCCGATCATCAACA	6840
Qy	1921	CYLeuLeuCYaAaenCYaSerTYrProThrArgLYSerLeuSerArgValCYaAaenProGlnAaP	Qy	2281	ProValValSerILESerValHISaAaenPheArgLYglnLeuLeuProArgLYaLeuAaenThr	2300
Db	5761	TGCTCTGTGTGTGACTGTACCCCAAGGCTCTGTGTCCAGAGTGTGTGACCTTGAGAT	Db	6841	CCCGTGTGAGCATCAAGGCTCATGATGATGAGAGCTTGTGCGCCGGGCGCTGGACAA	6900
Qy	1941	GLYglnCYaProCYaLYaPProGLYValILEGLYArgGlnCYaAaenArgCYaAaenPro	Qy	2301	ProValThrValGlnPheArgLYeLeuGlnThrGlnGlnArgTYrThLYeProILECYaVal	2320
Db	5821	GGCCAGTGTCCATGACAGGCAAGCTGTACACGGGCGGTCAAGTGTGACCGCTGTACAACTT	Db	6901	CCCGTCAAGGTGACGTTCCGCTGTGGAGACAGAGGCGGACCAAGCCCATCTGTGTC	6960
Qy	1961	PheAlaGlnValThrThrAaenGLYCYaGLYValAaenTYrAaenSerCYaProArgLYaILE	Qy	2321	PheThrAaenHISerILELeuValSerGLYThrGLYglnTYrPProSerAlaArgLYCYaSLu	2340
Db	5881	TTTGTGTGAGTCAACCAATGCTGTGTAAGTAATATGACAGGCTGCCCAAGACCAAT	Db	6961	TTCTGGAACCAATTCAATCTGTGTGAGTGCACAGGTGCTGTGGCCAGAGGCTGTGAA	7020
Qy	1981	GLuAlaGLYILETYrThrProArgThrArgPheGLYLeuProAlaAlaAlaProCYaPro	Qy	2341	ValValPheArgAaenGlnSerHISaSerCYaSLuCYaAaenHISmetThrSerPheAla	2360
Db	5941	CAGGCTTGGGATCTGGTGGCCGCTACCCGCTTCCGAGCTGCTGCTGCTCTGCTGCC	Db	7021	GTCGCTTCCGCAATGAGAGCACGTCAAGCTGCAGAGTCAACCAATGACAGCTTGCT	7080
Qy	2001	LYeGLYSerPheGLYThrAlaValArgHISCYaAaenGlnHISArgGLYrTYrLeuProPro	Qy	2361	ValLeuMetAaenValSerArgArgLYaAaenGLYglnILEuProLeuLYeThrLeuThr	2380
Db	6001	AAAGGCTCTTTGGGACGTCTGCGGCCACTGTGATGAGCAAGGGGGTGGCTCCGCCA	Db	7081	GTCGTCATGAGACGTTTCTCCGCGGGAGAAATGGGAGATCTTGCCATGAAACACTGACA	7140
Qy	2021	AaenLeuPheAaenCYeThSerILEThrPheSerGLYLeuLYeGLYPheAlaGlnArgLeu	Qy	2381	TYrValAlaLeuGLYValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeu	2400
Db	6061	AACTCTTCAACTGACGTCACATCACTCTCTCAGAACTGAAAGGCTTCCCTGACGGCTCA	Db	7141	TACGTGGCTTGAAGTGTCACTTGCTGCTGCCCTTGTGCTCACTTCTTCTCTCACTCTC	7200
Qy	2041	GlnArgAaenGlnSerGLYLeuAaenSerGLYArgSerGlnGlnLeuAlaLeuLeuLeuArg	Qy	2401	LeuArgLYLeuAaenSerAaenGlnHISgLYILEaArgArgAaenLeuThrAlaAlaLeuGLY	2420
Db	6121	CAGCGGAATGAGTCAAGGCTTACGACCTCAAGGCGCTCCCAAGCACTAGCCCTGCTGCGC	Db	7201	TTGGGTATCTTGCGCTCCCAACCAACAGGCAATCCGAGCTGACCTGACAGCTGCTGGGC	7260
Qy	2061	AaenAlaThrGlnHISThrAlaGLYTYrPheGLYSerAaenValLYeValAlaArgLYLeu	Qy	2421	LeuAlaGlnLeuValPheLeuLeuGLYILEaenGlnAlaAaenLeuProPheAlaCYeThr	2440
Db	6181	AACGCCACGACAGACACAGCTGGCTACTTCGCGACAGGACGTCAAGSTGGCTTACCAAGCTG	Db	7261	CTGGCTCAAGCTGCTTCTCTCTGGAATCAACAGGCTGACCTTCTTGTGCTGACACA	7320
Qy	2081	AlaThrArgLYeLeuAlaHISglnSerThrGlnArgLYPheGLYLeuSerAlaThrGln	Qy	2441	ValILEaAlaILEuLeuHISpHeLeuTYrLeuCYeThrPheSerTYrPAlaLeuLeuGln	2460
Db	6241	GCCAGCGCGCTGCTGCCCAAGAGACCCAGCGGGCTTGTGGCTGTCTGCCACACAG	Db	7321	GTCATTTGCCATCTTGCTGACCTTCTGTACCTCTGACCTTCTTCTGCGGCTGCTGGAG	7380
Qy	2101	AaenValHISpHeThrGlnAaenLeuArgValGLYSerAlaLeuLeuAaenPheAlaAaen	Qy	2461	AlaLeuHISLeuTYrArgAlaLeuThrGlnValArgAaenValAaenThrLYeProMetArg	2480
Db	6301	GACGTGCACTTCACTGAGAAATCTGCTCGGAGTGGGAGCGGCTCTCTGACACACAGC	Db	7381	GCTTGGACCTGTACCGGCACTCACTGAGGTGGCAATGTCAACACCGGCGCCCATGGGC	7440
Qy	2121	LYaArgHISrTYrGLYLeuLISglnGlnThrGlnGLYgLYThrAlaArgLYeLeuGlnHIS	Qy	2481	PheTYrTYrMetLeuGLYrTYrGLYValProAlaPheILEThrGLYLeuAlaValGLYLeu	2500
Db	6361	AAGCGGCACTGGGAGCTGATCCAGACAGACAGAGGTGGCAACGCTGCTGTCTCAAGCAC	Db	7441	TTCTACTACATGCTGGGCTGGGCGGTGCTGCTTCAATCAAGGCGCTAGCGGTGGCGCTG	7500
Qy	2141	TYrGLYAlaTYrAlaSerAlaLeuAlaGlnAaenMetArgHISThrTYrLeuSerProPhe	Qy	2501	AaenProGlnGLYTYrGLYaAaenProAaenPheCYeThrLeuSerILETYrAaenThrLeuILE	2520
Db	6421	TATGAGGCTTACGCGAGTGCCTTGGCCAGAACATGCGGCAACCTTACTTAAGCCCTTTC	Db	7501	GACCCCAAGGCTTACGAGAACCTTGACTTGTGCTGCTCTCCCACTTATGACACCTCATTC	7560
Qy	2161	ThrILEValThrProAaenILEValILESerValValArgLeuAaenPheGLYaAaenPheAla	Qy	2521	TYrSerPheAlaGLYProValAlaPheAlaValSerMetSerValPheLeuTYrILEu	2540
			Db	7561	TGAGATTGTGTCGCGCGGCTGGCTTGGCGCTTGCCTGATGAGTGTCTTCTTACATCTTG	7620

QY 2541 AlAlaArgAlaSerCyaaAlaAlaArgGlnGlyPheGluYsYsGlyProValSer 2560  
Db 7621 GCGGCGCGGCGCTCTCTGCTGCGCCAGCGGCGGCTTTAGAGAAAGAGTCTGCTCG 7680  
QY 2561 GYLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeu 2580  
Db 7681 GCGCTGAGCGCTCTCTGCGCTCTCTGCTGCTGAGCGCCAGTGGCTGCGGACCTG 7740  
QY 2581 LeuSerValaSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnGlyLe 2600  
Db 7741 CTCTCTCTCAACGCGACACCTCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCT 7800  
QY 2601 GlnGlyProPheLeuPheLeuSerTyrValValLeuSerGlyValAlaGlyAlaLeu 2620  
Db 7801 CAGGCGCGCTTCACT 7860  
QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
Db 7861 AACCTTGGCTGCGAGCGCCAGCGCCAGCGCTGAGCTGCTGAGCCAGCTCACCTCTG 7920  
QY 2641 ThrSerSerTyrAsnGlyProSerProTyrAlaAspGlyValGlyLeuTyrGlnProTyrGly 2660  
Db 7921 ACCTCTGCTCAACAATGCCAGCCCTTACGACAGATGGCGGCTGACAGCCCTTACGGA 7980  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrTle 2680  
Db 7981 GACTCGGCGCGCTCTCTGACACACACAGTGGCTCGGCGAAGTCCAGCCAGCTACATC 8040  
QY 2681 ProPheLeuLeuArgGlnGlnGlnSerAlaLeuAspProGlyGlnGlyProPogLysLeuGly 2700  
Db 8041 CCTCTCTGCTGAGGAGGAGGCTCGACCTGACCTGCGCCAGCGCGCTGCGCGG 8100  
QY 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
Db 8101 GATCCAGGCGCGCTGCTCTGAGAAAGTCAAGACCAAGACAGATGCTCTGACAGACTCC 8160  
QY 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
Db 8161 GACAGTGAAGCTGCTCTTGAAGAGAGACAGAGTGGCTCTCTATGCTCTACCTCATCA 8220  
QY 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
Db 8221 GACAGTGAAG 8280  
QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGlnArgLeuProLeuHisSerThrProLysAsp 2780  
Db 8281 TGGGATAGCTGCTGCGGCGCTGAGGAGAGAGACTGCTGCTGACAGTATCTCCCAAGAT 8340  
QY 2781 GlnGlyProGlyProGlyLysAlaProTyrProGlyAspPheGlyThrThrAlaLysGln 2800  
Db 8341 GGGGGCGCCAGGCGCTGCGAGAGGCGCCCTGCGCAGAGACTTTGGAGCCACAGCAAAAG 8400  
QY 2801 SerSerGlyAsnGlyAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2820  
Db 8401 AGTAGTGGCAACGGGCGCTCTGAGAGAGGCGCTGCGGAGATGAGATGAGGCTGCTCGA 8460  
QY 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisYsGly 2837  
Db 8461 GAGGGGCTCTTCAAGGCGCTCTTCCAGGCTCTTCCAGGCTCAAGCTCAAGAT 8511

RESULT 4  
US-10-292-798-931  
; Sequence 931, Application US/10292798  
; Publication No. US2003023583A1  
; GENERAL INFORMATION:  
; APPLICANT: SUMA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ASURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798

;; CURRENT FILING DATE: 2002-11-13  
;; PRIOR APPLICATION NUMBER: 10/017,161  
;; PRIOR FILING DATE: 2001-12-18  
;; PRIOR APPLICATION NUMBER: JP 2001-246789  
;; NUMBER OF SEQ ID NOS: 2070  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 931  
;; LENGTH: 24370  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; LOCATION: source  
;; FEATURE:  
;; LOCATION: (1) .. (24370)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (201) .. (3510)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (8553) .. (9200)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (11163) .. (11385)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (11634) .. (11749)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (11929) .. (12017)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (12408) .. (12565)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (12927) .. (13090)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (13263) .. (13388)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (13732) .. (13898)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (14199) .. (14486)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (14575) .. (14749)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (14989) .. (15130)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (15312) .. (15432)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (15856) .. (16039)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (16226) .. (16327)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (17669) .. (17763)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (17972) .. (18181)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (18702) .. (18885)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (19001) .. (19129)  
;; FEATURE:

```

NAME/KEY: CDS
LOCATION: (19231) .. (19411)
FEATURE:
NAME/KEY: CDS
LOCATION: (19545) .. (19712)
FEATURE:
NAME/KEY: CDS
LOCATION: (19815) .. (19941)
FEATURE:
NAME/KEY: CDS
LOCATION: (20054) .. (20254)
FEATURE:
NAME/KEY: CDS
LOCATION: (20547) .. (20721)
FEATURE:
NAME/KEY: CDS
LOCATION: (21048) .. (21160)
FEATURE:
NAME/KEY: CDS
LOCATION: (21338) .. (21417)
FEATURE:
NAME/KEY: CDS
LOCATION: (21507) .. (21633)
FEATURE:
NAME/KEY: CDS
LOCATION: (21721) .. (21843)
FEATURE:
NAME/KEY: CDS
LOCATION: (22339) .. (22526)
FEATURE:
NAME/KEY: CDS
LOCATION: (22761) .. (22849)
FEATURE:
NAME/KEY: CDS
LOCATION: (22954) .. (23148)
FEATURE:
NAME/KEY: CDS
LOCATION: (23287) .. (23457)
FEATURE:
NAME/KEY: CDS
LOCATION: (23557) .. (23791)
FEATURE:
NAME/KEY: CDS
LOCATION: (24143) .. (24170)
US-10-292-798-931

Alignment Scores:
Pred. No.: 0
Score: 1103.00
Percent Similarity: 100.00%
Best Local Similarity: 37.74%
Query Match: 12
DB: 0

Length: 24370
Matches: 1103
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-916-849a-3 (1-2923) x US-10-292-798-931 (1-24370)
QY 1 MetATGserProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
DB 201 ATGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 260
QY 21 LeuLeuLeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40
DB 261 TTGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320
QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSer 60
DB 321 GGGTCCAGGGGAGAGAGGCTCTTCGGGGGCGCTGCGCCCAATGGGCTGCTGCTCATCC 380
QY 61 SerAlaSerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThr 80
DB 381 TAGGGGTCCAGACCTGCTGCTCTACACAGCCGCTGCGAGGATGCGGCGACATGAGCTGACT 440
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValThrCysProGluSerGluHis 100

```

```

DB 441 GGCCACCTGTATCCCAACAGATGCTGTAGGGTTGTGTCCAGATCCGAGGCCCAT 500
QY 11eProLeuProProAlaProGluGlyCysProThrPserCysArgLeuLeuGlyValGly 120
DB 501 ATTCCTTACCAACAGAGCTCTGAAAGCTGCCCCCTGTAGCTGTGCTTGGGCAATGGA 560
QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGluHisProCysLeuValAla 140
DB 561 GGCCACCTTCCCAACAGGGAGGTCACTGCCCCGAGAGACCCGTGCTTAAGGCT 620
QY 141 ProArgLeuArgCysGlnSerCysAlaLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160
DB 621 CCACGGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
QY 161 ArgSerProGluGluSerLeuGlyGlyValArgArgValArgAsnValAsnThrAlaProGln 180
DB 681 AGGTCCAGAGAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
QY 181 PheGlnProProSerGlyGlnAlaThrValProGluHisGlnProAlaGlyThrProVal 200
DB 741 TTCCAGCCCCCAAGCTACAGGCCAGTCCGAGAGACAGCCAGCCAGCCCTGTT 800
QY 201 AlAserLeuArgAlaAlaAspProAspGlyGlyAlaGlyValGlyLeuGlyThrMet 220
DB 801 GCATCCCTGAGGGCCATGACCCGAGACAGGGGTGAGGGGTGACCTGAGCTACACCATG 860
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
DB 861 GATGCCCTCTTGTATAGCGCTCCCAACCGATCTCTCTCCGAGCCAGCTACCTGATCA 920
QY 241 ValThrThrAlaGluGluLeuAspArgGlyLeuThrLysSerThrHisValPheArgValThr 260
DB 921 GTAACCAAGCCGAGAGAGCTGATGATGAGACCAAGAGCCACCGTCTTACGGGTCAAG 980
QY 261 AlAsnLeuAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrLysLeuValThr 280
DB 981 GCGCAGGACCAAGGAGATGCCCCGAGAGAGTCCCTGGGTACACCTCATCTGATGTTACT 1040
QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGlyValGlySerLeuAspArgGly 300
DB 1041 GACACCAATGACCAATGACCTCTGTGTCGACGAGAGTACAGAGAGCTCCAGGAG 1100
QY 301 AsnLeuGluValGlyValGluValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro 320
DB 1101 AACCTGAGGTGGCTATGAGGTGCTCATCTGACAGGCGCCAGATGAGATGCCCTCC 1160
QY 321 AsnAlaAsnLysLeuThrArgLeuLeuGlyGlySerGlySerProSerGlyValPhe 340
DB 1161 AATGCAATATTCGTACCGCTCTGAGAGGGGTCTGGGGGACAGCCCTCTGAAGTCTTT 1220
QY 341 GluLysAspProArgSerGlyValLysArgThrArgGlyProValAspArgGlyGluVal 360
DB 1221 GAGATGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280
QY 361 GluSerTyGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380
DB 1281 GAATCTTACAGCTGACGGTGAAGGACCAAGTACAGAGGATCCGGGATCCCTCGAGAGT 1340
QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAsnAlaProGluPheSer 400
DB 1341 ACCACAGCGCTGTTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1400
QY 401 GluLysArgTyValAlaGlnValAlaArgGluAspValThrProGlyAlaProValLeuArg 420
DB 1401 GAGAGAGGCTATGATGCTCAGGTGAGGAGATGATGATCCAGGGGCGCCAGTACTCGGA 1460
QY 421 ValThrAlaSerAspAlaGlyGlySerAsnAlaValAlaHisTySerLysMetSer 440
DB 1461 GTCAAGGCTGATGATGAGACAGAGGAGGACATCCGCTGCTGCTGCTGCTGCTGCTGCT 1520
QY 441 GlyAsnAlaArgGlyGlnPheTyLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460

```

Db 1521 GGCAATGCTCGGGACAGTTTATCTGATGCCAGACTGAGCTCTGATGTGTAGC 1580  
 Qy 461 ProLeuaspTyrGluThrThrylsGluTyrThrLeuargValArglaaglNaapGly 480  
 Db 1581 CCTCTGACATAGAGACAGCAAGAGTACACCTTACGGGTGGAGACACAGATGGTGGC 1640  
 Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp1LeaAsp 500  
 Db 1641 CGTCCCCCACTCTATATGTCTCTGGCTGTGTGACAGTACAGGTCTCTGGATATCAACGAC 1700  
 Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 Db 1701 AATGCCCCCATCTTCGACAGACCCCTTTCAGGCTACGTCTCGAGAGTGTCCCTTA 1760  
 Qy 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db 1761 GGTACTCTGGTCTTCATGTCCAGGCTATGACGCTGATGCTGTGTACAAATGCCCTGTG 1820  
 Qy 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db 1821 GAATACCGCTCTGCTGGGTGGAGCATGACTCCCTTCAACATCAATGACACAGGC 1880  
 Qy 561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
 Db 1881 TGGATCTCTGTGGCTGTGTAAGTGAACCGGAGAGAGTATTTCTACAGCTTTGGAGTA 1940  
 Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1941 GAAAGCTCGAAGCATGGCACTCCAGCATCTACGTCCCTGGCCAGTGTGACGCTGACTGTG 2000  
 Qy 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValAlaArgLeuAsn 620  
 Db 2001 CTGGAAGTCAACACACAAATCCAACTTACCAACAGATACACAGTACAGTGGCGCTCAAT 2060  
 Qy 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspAspAlaHis 640  
 Db 2061 GAGATGACACTGTGGACACAGGAGTGTGACGGTGTGACGTGTGGACCGTATGCTCAT 2120  
 Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 2121 AGTGTATCACTACCAAGATCAACAGTGGCAATATCGAAGACCGCTTCTCATCACAGC 2180  
 Qy 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln 680  
 Db 2181 CAAGAGGTGTGGCTGTATCCCTTGTCCCTGTGCACTGACATCAAACTTGAGCGGACG 2240  
 Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 700  
 Db 2241 TATGTGTGCTGTACCGCTCCGATGTGCACTGCGACAGACAGGCAAGATTGTGTG 2300  
 Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisIleThrValAsn 720  
 Db 2301 AATGTACCGACCGCAACCCCATCGTCTCTTCAAGAGCTCCCATATACAGTGAAT 2360  
 Qy 721 ValAsnGlnAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlnAsp 740  
 Db 2361 GTTAAATAGAGACCGCGGACGACCAACGCTGTGATGACGCGCAAGATGAGAC 2420  
 Qy 741 ThrGlyLeuAsnAlaArgIleThrTyrPheMetGlnAspSerIleProGlnPheArgIle 760  
 Db 2421 ACAGGTGAGATGCCCGCATCACTTCTATGAGAGACAGCATCCCGCATTCGCAATC 2480  
 Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGlnAspGlnValSer 780  
 Db 2481 GATGCAACACAGGGGGCTGTCAACCAAGCTGAGCTGATATGAAAGCAAAAGTGTCT 2540  
 Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrThr 800  
 Db 2541 TACACCTGTGGCATTAATCTCGGAGCAAAATGCAATCCCAAGAGTCCGACCACTTAC 2600  
 Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyr 820  
 Db 2601 CTGGAAGATCTCTGGTGAACGAGATGACATGCAATGCCCTCAAGTTCTCTGCGAGACTCTTAC 2660

Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2661 CAGGCAAGTGTATAGAGATGTGCAACCTTCACTAGAGTCTCTCAAGTCTCAAGCACT 2720  
 Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2721 GATCTGTATTTCTGACCTTAATGTGCAAGGTCCTTCACTCACTTCCAAAGGCGCACATGGA 2780  
 Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 Db 2781 GACGTTGACTTATTTATGTGATCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2840  
 Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPro 900  
 Db 2841 CGAGAGAGCTGGCCCAATGATGTCTGGGGGCAATGACAGTGGACAGAGGATGCCCA 2900  
 Qy 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspProVal 920  
 Db 2901 GCCCGCACACTATGGAAGTGAAGTCACTGTGTGTGATGTGAATGACATATCCCTGTTC 2960  
 Qy 921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 Db 2961 TTGAGCAGATGATGATGTGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3020  
 Qy 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 3021 GCCCGGTCAAGCAACATGACCCCGATGAGGCAACATATGCCAATATATGATACAGATT 3080  
 Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla 980  
 Db 3081 GTGGAGGGCAATCATCTGAGAGTCTTTCAGCTGACATCTTCTCCGGGAGCTGACAGCC 3140  
 Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000  
 Db 3141 CTGTGAGACTTGACTTACAGAGACCGGCTGAGTACGTCTGTATCCAGGCAAGTGA 3200  
 Qy 1001 AlaProLeuValSerArgAlaThrValHisValAlaArgLeuLeuAspArgAsnAspAspPro 1020  
 Db 3201 GCTCCTGTGTGAGCGGGCTCAAGTCAAGTCCCTCTTGAACGCGCAATCAACCA 3260  
 Qy 1021 ProValIleLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040  
 Db 3261 CCAAGTCTGGGCAACTTGAATCTTTCACCAACTATGTACCAATGCAATGCTCAAGCAAC 3320  
 Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 3321 TTCCCTGGGGGTGCCATTTGGCCGAGTACTGCTCCATGACCTGATATCTCAATGTGTG 3380  
 Qy 1061 ThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 Db 3381 ACTTACAGCTTGAACGGGGGAATGAACCACTCGTGTCTCTCATATGCTTCCACGGGT 3440  
 Qy 1081 GluLeuValSerSerArgAlaLeuAspAsnAsnAspProLeuGlnAlaIleMetSerVal 1100  
 Db 3441 GAGCTGAAGCTTAAGCGCGCATGACAAACCGGCTCTGAGGCGCATATGAGCGTG 3500  
 Qy 1101 LeuValSer 1103  
 Db 3501 CTGGTGTCA 3509

RESULT 5  
 US-09-764-870-569  
 ; Sequence 569, Application US/09764870  
 ; Patent No. US20020042386A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT214  
 ; CURRENT APPLICATION NUMBER: US/09/764,870  
 ; PRIORITY FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 646

SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 569  
LENGTH: 2332  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-870-569

## Alignment Scores:

Pred. No.:	0	Length:	2332
Score:	776.00	Matches:	776
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	26.55%	Indels:	0
DB:	9	Gaps:	0

US-09-916-849A-3 (1-2923) x US-09-764-870-569 (1-2332)

QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
DB 3 ACCCAAGCTTTCAGGGGTCAAGGCGAGACCAAGGCGATGCCCGACGAAGTGCCTGGCT 62  
QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlu 293  
DB 63 ACACTCAACATCTTGTTATCTGACACCAATGACATGACCTGTGTTCGAGCAGCAGAG 122  
QY 294 TyrIleGlnSerLeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValAlaAla 313  
DB 123 TACAAGGAGAGCTCAGGAGAGAACTGAGAGTTGGCTATGAGTGTCTCACTGCAGGGCC 182  
QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGly 333  
DB 183 ACGAGTGTGATGCCCCCTCCCAATGCAATATCTGTACCGCTGTGAGGGGTCTGGG 242  
QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353  
DB 243 GGGAGCCCTCTGAAGTCTTGAATGACCTCTCTGGGGTATCCGAACCCGTGGC 302  
QY 354 ProValAspArgGluGluValGlnSerTyrGlnLeuThrValGluAlaSerAspGlnGly 373  
DB 303 CCTGTGATCCGGAGAGAGGTGAACTCAACAGCTGACGGTGAAGCAAGTACAGAGGT 362  
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsn 393  
DB 363 CGGAGCCGGGCTCTCGAGATCAACAGCCGCTGTTTCTTCTGTGAGAGATACAT 422  
QY 394 AspAsnAlaProGlnPheSerGluIleAspGlyValIleGlnValArgIleAspValThr 413  
DB 423 GATATGCCCCCAATTAATGTAAGAACCCCTATGTGTCAGGTGAGGAGATGTGACT 482  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
DB 483 CCAGGGGCCCCAGTACTCCAGTCAACAGCTCGATCCAGACAAAGGAGCAATCCGTG 542  
QY 434 ValHisTyrSerIleMetSerGlyValAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
DB 543 GGGCACTATAGCATATGATGATGATGCTCGGGACAGTTTATCTGATGATCCAGACT 602  
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGlyTyrThrLeuArg 473  
DB 603 GGAAGCTCTGAGATGTGTGAGCCCTCTTGACTATGAGACCAAGAGATACCCCTTACG 662  
QY 474 ValAlaGlnIleAspGlyValArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
DB 663 GGGCAGACAGAGAGTGTGCGCTCCCACTCTAATGTCTCGGCTGTGTGACATTA 722  
QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
DB 723 CAGGTCTCGAATATCAACACATGCCCCCATCTTGTCAGACACCCCTTCCAGGCTACT 782  
QY 514 ValIleGlnSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
DB 783 GTCTCGAGAGAGGTCCCTTATGAGCTACCTGTTCTCAATGTCTCAAGCTATGACCCCTAT 842

QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValIleAspPheProPhe 553  
DB 843 GCTGTGACAAATGCCCCGCTGGAATATACCGCTTCTGGGGTGGAGACATGATCCCTTCC 902  
QY 554 ThrIleAsnGlnGlyThrIleThrIleSerValAlaAlaGluLeuAspArgGluValIle 573  
DB 903 ACCATCAACAATGACACAGGCTGATCTGTGTCTGTCTGTCACTGACCGGAGAGATTT 962  
QY 574 AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
DB 963 GATTTCTACAGCTTTGGGGTGAAGCTCGACCAATGCACTCCAGCACTACCTCCCTCG 1022  
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613  
DB 1023 GCCAGTCAAGCGTGAAGTGTCTGTGATGTCAACAGACAAATCAACTTATACCAACCA 1082  
QY 614 GluTyrThrValAlaGlnLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer 633  
DB 1083 GAGTACACAGTGCAGCTCAATGAGAGATGCAAGTGTGGACACAGCGTGTACGCTGCA 1142  
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653  
DB 1143 GCTGTGACCTGTATGCTCATATGTTATCTATCACTACCAATGACACAGTGGCAATCTGCA 1202  
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyIleLeuValSerLeuAlaLeuProLeu 673  
DB 1203 AACGCTTCTTCATCACACGCAAGAGTGTGGTGGCTGTGATCCCTTGCCTGCCACTG 1262  
QY 674 AspTyrIleLeuGluValArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
DB 1263 GACTCAAACTTGAAGCGGAGATATGTGTGTGTGTACCGCTCCAGTGCATCGGAG 1322  
QY 694 AspThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisAspProValPheGln 713  
DB 1323 GACACGGCAGAGATGTGTGTATGTCAACGACCCAAACCATGCTGTCTTTCAG 1382  
QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrValValLeu 733  
DB 1383 AGCTCCCACTATACAGGAATGTTAATGAGACCGCGCGGACACACCGTGTGTCTG 1442  
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753  
DB 1443 ATCAAGCCGACAGATGAGAGACAGTGTAGATCCCGCATCACTTATATGAGAGAC 1502  
QY 754 SerIleProGlnPheAspGlyLeuAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu 773  
DB 1503 AGCATCCCCCACTCCCATGATGACGACACGSGGCTGTCAACCCAGGCTGAGCTG 1562  
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyTyrPro 793  
DB 1563 GACTACGAAGACCAAGTGTCTTACACCTGTGCCATTTACTGCTCGGACAAATGGCATTCCTC 1622  
QY 794 GlnIleSerAspThrThrTyrIleGluGlnIleValIleAspValIleAspAsnAlaPro 813  
DB 1623 CAGAACTCCACACACACTTACCTGTGAGATCTGTGTGAACACAGGAATGAAATGATGCCCT 1682  
QY 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833  
DB 1683 CAGTTCTCGAGAGCTCTTACACAGGCGAGTGTCTATAGAGATGCGCACCTTACATGAC 1742  
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyValArgValPheTyrThr 853  
DB 1743 GTCTCGAGATCTCAGCAGCTGATCGTGAATTTCTGATATGCAAGGGTCTTCTTACACC 1802  
QY 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyTyrVal 873  
DB 1803 TTCCAAAGAGCGACATGAGAGAGTGAATTTATGTGTGATGATCAAGTCAAGGATGCTG 1862  
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla 893  
DB 1863 CGAAGCTACGAGAGGCTGATGATGAGAGAGACGTGGCCAGATATGTCTTGGCGGACATATGCA 1922  
QY 894 ValAspIleGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913

DB 1923 GTGACAGAGGATGCCCCAGCCCGACCACTATGGAAGTGAAGTCACTGTGTGGAT 1982  
QY 914 VALAAspAspAspProProValPheGluGlnAspGluPheAspValPheValGluGluAsn 933  
DB 1983 GTGAAGACAAATCCCCCTGTCTTGAAGAGATGATGTGTGTGTGTGAAGAAAC 2042  
QY 934 SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn 953  
DB 2043 AGCCCATTTGGGTACCGCCGTGGCCCGGTCAAGCCATCAACCCCATGAAGCAACAT 2102  
QY 954 AlaGlnIleMetSerGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIle 973  
DB 2103 GCCCAGATTATGATACAGAGATTGTGAAGGCAATCCCTGAGTCTTCCAGCTGAGCATC 2162  
QY 974 PheSerGlyGluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyrVal 993  
DB 2163 TTCTCCGGGAGCTGACAGCCCTGTGAGACTGACAGAGACCGGCTGTAGTACGTC 2222  
QY 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1013  
DB 2223 CTGGTATCCAGGCCACGTACGCTCTGTGTAGCCGGCTCAAGTCCAGTCCGCTTC 2282  
QY 1014 LeuAspArgAspAspAspProProValLeuGlyAsnPheGlnIleLeu 1029  
DB 2283 CTTGACCGCATGACAAACCAAGCTGTGGCACTTGAGATCCTT 2330  
RESULT 6  
US-10-125-540-569  
/ Sequence 569, Application US/10125540  
/ Publication No. US20030059875A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Rosen et al.  
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
/ FILE REFERENCE: FT214C1  
/ CURRENT APPLICATION NUMBER: US/10/125,540  
/ PRIOR FILING DATE: 2002-04-19  
/ Prior Application removed - See File Wrapper or Palm  
/ NUMBER OF SEQ ID NOS: 646  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 569  
/ LENGTH: 2332  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-10-125-540-569  
Alignment Scores:  
Pred. No.: 0 Length: 2332  
Score: 776.00 Matches: 776  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.55% Indels: 0  
DB: 15 Gaps: 0  
US-09-916-849a-3 (1-2923) x US-10-125-540-569 (1-2332)  
QY 254 ThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAla 273  
DB 3 ACCCACTCTTCAGGGTCAAGCGGAGACCAAGGATGCCCGGAGCAATGCTGTGGCT 62  
QY 274 ThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlu 293  
DB 63 ACACTCAACCATCTTGTTACTGACCAACATGACCATGACCTGTGTGTGAGCAGAGAG 122  
QY 294 TyrIleGluSerIleuValGluLeuGluValGlyTyrGluValLeuThrValArgAla 313  
DB 123 TACAGAGAGAGCTCAGAGAGAACTGAGGTTGCTATAGAGGTCTCACTGTCAAGGCC 182  
QY 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuGluGlySerGly 333  
DB 183 ACGGATGTATAGCCCTCCCAATGCCAATTTCTGTATACCGCTGTGAGAGGGGTCTGG 242  
QY 334 GlySerProSerGluValPheGluIleAspProArgSerGlyValIleArgThrArgGly 353

DB 243 GGCAGCCCTCTGAAGTCTTTGAGATGACCTCGCTCTGGGGTATGCCAACCCTGGC 302  
QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnIle 373  
DB 303 CCTGTGATCGGGAGAGAGGTGAATCTTACCAAGCTGAGAGGTGAGAGCAAGCAAGGT 362  
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393  
DB 363 CGGAGCCCGGCTCTCGAGATGCCAGCCCGCTGTTCCTTCTGTGTGAGATGACAT 422  
QY 394 AspAsnAlaProGlnPheSerGlyLeuArgTyrValValGlnValArgGluAspValThr 413  
DB 423 GATATATGCCCCCAGATTATGAGAAAGGCTATGTGTCCAGATGAGAGAGATGTGACT 482  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAspAlaVal 433  
DB 483 CCAAGGGGCCAGATCTCCAGATCAAGCTCGATTCAGACAGAGGAGAGCAATGCCGTG 542  
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453  
DB 543 GTGCATATAGCATATATAGTGGCAATGCTCGGGACAGTTTATCTGGATGCCAGACT 602  
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGluTyrThrLeuArg 473  
DB 603 GGAAGCTGTGATGTGTAGCCCTCTTGAATATAGAGACCAAGAGATGACCTTACG 662  
QY 474 ValArgAlaGlnAspGlyValArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
DB 663 GTGCAGAGACAGAGAGTGGCCGCTCCCACTCTTAATGTCTGTGGTGGTGAAGATA 722  
QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
DB 723 CAGGTCTGTGATATCAAGACATGCCCTTCCTGTGTCAAGACCCCTTCCAGGCTACT 782  
QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
DB 783 GTCCTGAGAGAGGCTCCCTTGAAGCTAGCTGTCTCCATGTCCAGGCTATGACCTCAT 842  
QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValAlaHisAspPheProPhe 553  
DB 843 GCTGTGACAAATGCCCGCTGTGAATACCGCTGTGTGGGTGGAGACATGATCTCCCTTC 902  
QY 554 ThrIleAsnAsnGlyThrGlyTyrIleSerValAlaAlaGluLeuAspArgGluGluVal 573  
DB 903 ACCATCAACATAGGACAGAGGCTGATCTGTGTGTGTGAACTGACCGGAGAGAGTT 962  
QY 574 AspPheTyrSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
DB 963 GATTTCTACAGCTTGGGGTGAAGACTGAGACCAATGCACTCCAGCACTGCTCCG 1022  
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnPro 613  
DB 1023 GCGAGGTGACGGTACGTCTGTGATGTCAACGACAAATCAACTCTTACCCACCA 1082  
QY 614 GlnTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer 633  
DB 1083 GAGTACACAGTCCGCTCATATAGAGATGAGTGTGGGACACAGGTGGTGAAGGTGCA 1142  
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653  
DB 1143 GCTGTGACCGGTGATCATGTGTATCACTTCAACATCAACAGATCAAGAGGCAATCTGCA 1202  
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyValLeuValSerIleAlaLeuProLeu 673  
DB 1203 AACCGCTTCTCATATCAACCAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1262  
QY 674 AspTyrTyrLeuGluArgGlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
DB 1263 GACTACAAACTTGAACCGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1322  
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713

Db 1323 GACACGGCACAGATTGCTGAATGTCACCGACCCCAACCCATGCTCTTCTTCAG 1382  
 Qy SerSerHisThrValaenValaenValaenValaenValaenValaenValaenValaen 714  
 Db 1383 AGCTCCACATATACAGTAATGTTAATGAGACCGGCGGCGGACCAAGGAGGCTG 1442  
 Qy 115SerAlaThrProGluuThrGluuThrGluuThrGluuThrGluuThrGluuThrGluu 734  
 Db 1443 ATCAGGCGCCACGATGAGACACAGGAGTGCAGATCCGACATCACTTCATGAGAGAC 1502  
 Qy Ser1LeuProGlnPheArg1LeuAlaAspThrGluValaValaThrThrGlnAlaGluLeu 773  
 Db 1503 AGCATCCCGCCAGTCCCGCATCGATCGACGACGCGGCGCTTCACACCCAGCTGAGCTG 1562  
 Qy AspTyrGluAspGlnValSerTyrThrLeuAla1LeuThrAlaAspAsnGly1LeuPro 793  
 Db 1563 GACTCAGAGACCAAGTGTCTTACACCTGCGCATTACTCTCGGGACATGCAATGCC 1622  
 Qy GlnTyrSerAspThrThrTyrLeuGlu1LeuValaAsnAspValaAsnAspAsnAlaPro 813  
 Db 1623 CAGAAAGTCCAGACCACTACCTGAGATCTGGTGAACGATGAAATGCAATGCCCCCT 1682  
 Qy GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValaProPheThrSer 833  
 Db 1683 CAGTTCCTGGAGACTCTTACAGGCGAGTGTCTTAAGATGTGCCACCTTCATCAGC 1742  
 Qy ValLeuGln1LeuSerAlaThrAspArgAspSerGlyLeuAsnGlyValaValaPheTyrThr 853  
 Db 1743 GTCTCAGACTCTCAGCACTGATCGTGAATTCGACTTAATGCGAGGGCTTCTACAC 1802  
 Qy PheGlnGly1GlyAspAspGlyAspAspGlyValaGly1GlySerThrSerGly1LeuVal 873  
 Db 1803 TTCAGAGAGGCGACCAATGAGACGGTGAATTTATGTTAGTCAAGTCAAGGCAATCGTG 1862  
 Qy ArgThrLeuArgThrLeuAspArgGluAsnVala1GlyTyrValaLeuArgAlaTyrAla 893  
 Db 1863 CGAAAGCTACGAGGCGTGAATCGAGAAAGTGTGCCCAAGTATCTTGGCGGCAATGCA 1922  
 Qy ValAspLeuGly1MetProProAlaArgThrProMetGluValaThrValaLeuAsp 913  
 Db 1923 GTGCAAGAGGGATGCCCCAGCGCCCACTATGGAAGTGAAGTCACTGCTGTTGAGT 1982  
 Qy ValaAspAsnProProValaPheGluGlnAspGluPheAspValaPheValaGluGluAsn 933  
 Db 1983 GTGAATGACATCCCTCTGTTGAGCAGATGATGATTTGATGTTTGTGAGAGAGAAC 2042  
 Qy SerPro1GlyLeuAlaVala1AlaArgVala1ThrAlaThrAspProAspGluGlyThrAsn 953  
 Db 2043 AGCCCATTTGGGCTTACCGTGGCGGGGTCAAGCCATGACCCCGATGAAGGACCAAT 2102  
 Qy AlaGln1MetTyrGln1LeuVala1GluGlyAsn1LeuProGluValaPheGlnLeuAsp1Le 973  
 Db 2103 GCCCAATTAATGACAGATTTGAGAGGCAACATCCCTGAGGGTCTTCCAGCTGAGCATC 2162  
 Qy PheSerGly1GluLeuThrAlaLeuValaAspLeuAspTyrGluAspArgProGluTyrVala 993  
 Db 2163 TTCTCCGGGAGGTGACAGCCCTGTGATGATTAAGTACGAGGACCGGCTGATGATCGTC 2222  
 Qy LeuVala1LeuAlaThrSerAlaProLeuValaSerArgAlaThrVala1sValaArgLeu 1013  
 Db 2223 CTGGTATTCAGGCGCCATCGTCAAGTCTCTGGTGAAGCGGCGGCTCAAGTCAAGTCCGCTC 2282  
 Qy 1014 LeuAspArgAsnAspAsnProProValaLeuGlyAsnPheGlu1Leu 1029  
 Db 2283 CTGTAACCGAATGACAAACCAACGATGCTGGGCACTTGAAGATCCTT 2330  
 RESULT 7  
 US-10-176-847-99  
 ; Sequence 99, Application US/10176847  
 ; Publication No. US20030068636A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Velby, Peter Ole  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST  
 ; TITLE OF INVENTION: AND OVARIAN CANCER  
 ; FILE REFERENCE: MRI-039  
 ; CURRENT APPLICATION NUMBER: US/10/176,847  
 ; CURRENT FILING DATE: 2002-06-21  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: FASTSeq for Windows Version 4.0  
 ; SEQ ID NO 99  
 ; LENGTH: 2391  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-176-847-99  
 Alignment Scores:  
 Pred. No.: 0 Length: 2391  
 Score: 729.00 Matches: 729  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 24.94% Indels: 0  
 DB: 15 Gaps: 0  
 US-09-916-849A-3 (1-22923) x US-10-176-847-99 (1-2391)  
 Qy 2186 ArgTyrGluAlaLeuArgGlyGluGlnProProAspLeuGluThrThrVala1LeuPro 2205  
 Db 19 CGCTACGAGGCGCTCGTGGGAGACGCCCCGAGACCTTGACAAACAGTCACTGCTGCT 78  
 Qy 2206 GluSerValaPheArgGluThrProProValaValaAspProAlaGly1ProGlyGluAlaGln 2225  
 Db 79 GAGTGTCTTCAAGAGAGAGCCCGCTGATGAGCGCCGAGGCGCGGAGAGGCGCCAG 138  
 Qy 2226 GluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGlyGluAla 2245  
 Db 139 GAGCCAGAGAGCTGCGACGCGGACAGGACGCGCACCCGAGCTGAGCGCAGGCT 198  
 Qy 2246 Vala1AspVala1LeuTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspPro 2265  
 Db 199 GTGGCAGCGTATATATCAACGACCTGCGCGGCTCACTGCTCAATATATGACCT 258  
 Qy 2266 AspTyrAspSerLeuArgValaProAspArgPro1LeuLeuAsnThrProValaValaSer1Le 2285  
 Db 259 GACAAAGCGCACTTAAAGATGCCCAAGCCCGATCAACACACCCGTGTGAGCATC 318  
 Qy 2286 SerValaHisAspAspGluGluLeuLeuProAlaLeuAspTyrProValaThrValaGln 2305  
 Db 319 AGCGTCAATGATGATGAGAGCTTCTGCGCGGCTGAGCAACCCGTCAAGGTGAG 378  
 Qy 2306 PheArgLeuLeuGluThrGluGluArgThrTyrPro1LeuValaPheThrAsnHisSer 2325  
 Db 379 TTCCGCTGTGAGACAGAGAGGAGGACCAAGCCCATCTGTGCTTCTGGAACCATCA 438  
 Qy 2326 1LeuLeuAlaSerGlyThrGlyGlyTyrSerAlaArgGlyCyGluValaValaPheArgAsn 2345  
 Db 439 ATCTGCTCAATGAGGACAGGAGTGTGCTGCGGCGGCTGAGAGTGTCTTCCGCAAT 498  
 Qy 2346 GluSerHisValaSerCyGlnCyAsnHisMetThrSerPheAlaValaLeuMetAspVala 2365  
 Db 499 GAGAGCCAGTCAAGCTGCGACATGCAACCAATGACAGCTTCTGCTGCTCATGAGCCTT 558  
 Qy 2366 SerArgArgGluAsnGlyGlu1LeuLeuProLeuTyrThrLeuThrTyrVala1AlaGluGly 2385  
 Db 559 TCTCGCGGAGAAATGGGAGATCTCTGCACTGAAGACATGACATGATGATGCTTAACT 618  
 Qy 2386 ValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArg1LeuArg 2405  
 Db 619 GTCACTTGGCTGCTTCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 678  
 Qy 2406 SerAsnGlnHisGly1LeuArgArgAsnLeuThrAlaAlaGluGlyLeuAlaGlnLeuVala 2425  
 Db 679 TCACCAACCAACGCGCATCCGATCACTGACAGCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCT 738  
 Qy 2426 PheLeuGluGly1LeuGlnAlaAspLeuProPheAlaCySerThrVala1Ala1LeuLeu 2445



Db 739 TTCCTCTGGGAATCAACGAGCTGACCTTCCTTTGCTGCAAGTACATTCCTGCTG 798  
 Qy 2446 LeuH1ePheLeuTyrLeuCystrPheSerTrpAlaLeuLeuGluAlaLeuH1eLeuTyr 2465  
 Db 799 CTGCACCTTCTGTAACCTCTGCACCTTTTCTGGGGCTCTGCTGGAGGCTTTCACCTGTAC 858  
 Qy 2466 ArgAlaLeuThrGluValaArgAspValaAsnThrGlyProMetArgPheTyrTyrMetLeu 2485  
 Db 859 CGGGCACTCACTGAGGGCCGAGATGTCACACCGGGCCCAATGGCTTCTACTACATGCTG 918  
 Qy 2486 GLYTTRGGLVValaProAlaPheIleThrGlyLeuAlaValaGlyLeuAspProGluGlyTyr 2505  
 Db 919 GGGCTGGGGCGCTGCTGCTTCATCAAGGGGCTAGCCGTGGGCTTGGACCCCGAGGGCTAC 978  
 Qy 2506 GlyAsnProAspPheCystrTrpLeuSerIleTyrAspThrLeuIleTrpSerPheAlaGly 2525  
 Db 979 GGGAACTCTGACTTCTGCTGGCTCTCATATGACACGCTCATCTGGAGTTTGTGCTGGC 1038  
 Qy 2526 ProValaAlaPheAlaValaSerMetSerValaPheLeuTyrIleLeuAlaAlaArgAlaSer 2545  
 Db 1039 CCGGTGGCTTTGCTCCGTGATGAGTGTCTTCTGTACATCTGGGCGGGCTGCTCC 1098  
 Qy 2546 CyvalAlaAlaGlnArgGlnGlyPheGluValaValaGlyProValaSerGlyLeuGlnProSer 2565  
 Db 1099 TGTGTGCCCCAGCGGCGAGGCTTGAAGAAAGTCTGTCTGTGGGCTGCAAGCTTCC 1158  
 Qy 2566 PheAlaValaLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValaAsnSer 2585  
 Db 1159 TTGGCGGTCTCTGCTGCTGAGACGCGACCTGCTGCTGCACTGCTCTGTGCAACAGC 1218  
 Qy 2586 AspThrLeuLeuPheH1eTyrIleuPheAlaThrCyAsnCyAlaGlnGlyProPheIle 2605  
 Db 1219 GACACCTCTCTTCCACTACCTCTTGTGCTACCTGCAATTGCAATCAAGGGCCCTTCATC 1278  
 Qy 2606 PheLeuSerTyrTrpAlaValaLeuSerLysGluValaArgValaAlaLeuLysLeuAlaCySer 2625  
 Db 1279 TTCTCTCTCATATGGCTTACAGAGAGTCCGGAAGCACTCAAGCTTGTCTGACAGC 1338  
 Qy 2626 ArgLysProSerProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsn 2645  
 Db 1339 CGCAAGCCAGCCCTGACCTGCTGTGACCAACCAAGTCCCTGACCTGTCTTCAACAGC 1398  
 Qy 2646 CyPProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySer 2665  
 Db 1399 TGCCCCAGCCCTTACGAGATGGCGGCTGTACAGCCCTTACGAGAACTCGCGGCTCT 1458  
 Qy 2666 LeuH1eSerThrSerArgSerGlyLysSerGlnProSerTyrIleProPheLeuAlaArg 2685  
 Db 1459 CTGCACAGCACAGTGTGCTGGGCGAAGTCAAGCCCAAGCTTCTTGTGTAAGG 1518  
 Qy 2686 GlnGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeu 2705  
 Db 1519 GAGGAGTCCGCACTGAACCTTGCCCAAGGCCCCCTGCTGGGGATCCAGGCAACCTCG 1578  
 Qy 2706 PheLeuGlnGlyGlnAspGlnGlnH1eAspProAspThrAspSerAspSerAspLeuSer 2725  
 Db 1579 TTCTCTGAAAGTCAAGACACGAGCATGATCTGACACGCACTCCACAGTGCCTGTCC 1638  
 Qy 2726 LeuGlnAspAspGlnSerArgLysTyrAlaSerThrH1eSerSerAspSerArgLysGln 2745  
 Db 1639 TTGAGAACACACAGATGGCTCTTATGCTTACCACTACACAGACAGTGAAGAGAA 1698  
 Qy 2746 GlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeu 2765  
 Db 1699 GAAAGAGAGAGAGAGAGAGAGGCGGCTTCCCTGAGAGAGAGGCTGGATACCTGTGCTG 1758  
 Qy 2766 GlyProGlyValaGluValaGluProLeuH1eSerThrProLysAspGlyGlyProGlyPro 2785  
 Db 1759 GGGCGCTGAGACAG 1818  
 Qy 2786 GlyLysAlaProTyrProGlyLysAspPheGlyThrThrAlaLysGlnSerSerGlyLysGly 2805  
 Db 1819 GGCAGAGCCCTCTGGCCAGAGACTTTGGAGCCACAGCAAAAGAGTATGGCAACGGG 1878

Qy 2806 AlaProGlnGlnArgLysLeuArgGlyAsnGlyAspAlaLeuSerArgGlyGlySerLeuGly 2825  
 Db 1879 GCGCTTGAAGAGAGCGGCTCGGAGAGATGAGATGCCCTGTCTGAGAGAGGCTCTTAAAGC 1938  
 Qy 2826 ProLeuProGlySerSerAlaGlnProH1eLysGlyIleLeuValaLysCyLeuPro 2845  
 Db 1939 CCGCTTCAAGGCTCTTGTGCTCCAGCTCAAAAGGCACTTAAAGAAAGTGTGTGCC 1998  
 Qy 2846 ThrLysSerGlnLysSerSerLeuLeuArgProLeuGlnGlnCyThrGlySerSer 2865  
 Db 1999 ACCATCAGGAG 2058  
 Qy 2866 ArgGlySerSerAlaSerGlnGlySerArgGlyGlyProProProArgProProArg 2885  
 Db 2059 CGGGCTCTCGCTGCTAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2118  
 Qy 2886 GlnSerLeuGlnGlnGlnLeuAsnGlyValaMetProIleAlaMetSerIleLysAlaGly 2905  
 Db 2119 CAGAGCTTCCAG 2178  
 Qy 2906 ThrValaAspGluAspSerSerGlySer 2914  
 Db 2179 ACGGTGATGAG 2205  
 RESULT 8  
 US-09-764-870-570/c  
 ; Sequence 570, Application US/09764870  
 ; Patent No. US20020042386A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PTZ14  
 ; CURRENT APPLICATION NUMBER: US/09/764,870  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 646  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 570  
 ; LENGTH: 2077  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-764-870-570  
 Alignment Scores:  
 Pred. No.: 0 Length: 2077  
 Score: 691.00 Matches: 691  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.64% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-916-849a-3 (1-2923) x US-09-764-870-570 (1-2077)  
 Qy 2254 ThrH1eValPheAspValaThrAlaGlnAspH1eGlyMetProArgArgSerAlaLeuAla 273  
 Db 2075 ACCCAAGCTTCAAGGAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2016  
 Qy 2774 ThrLeuThrIleLeuValaThrAspThrAsnAspH1eAspProValaPheGlnGlnGln 293  
 Db 2015 AACCTCAACATCTTGTGATGACACCAATGACCAATGACCTGTGTGAGAGAGAGAG 1956  
 Qy 294 TyrLysLysSerLeuArgLysAsnLeuGlnValaGlyTyrGlnValaLeuThrValaArgAla 313  
 Db 1955 TACAAAGAGAGAGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1896  
 Qy 314 ThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGly 333  
 Db 1895 ACGAGTGGAG 1836  
 Qy 334 GlySerProSerGlnValaPheGlnLysAspProArgSerGlyValaIleArgThrArgGly 353  
 Db 1835 GGCAGCCCTCTGAAGCTTTGAGATGACCTTCCCTCTGGGGTATTCGAAACCCGTGCT 1776

```

QY 354 ProValAspArgGluGluValGluSerTyrGlnLeuThrValGluIleAspArgGlnGly 373
DB 1775 CCGTGGATCGGGAGAGAGGTGAAATCTACAGACTGACGGTAGAGCAAGTGCACGGGT 1716
QY 374 ArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393
DB 1715 CCGAGACCCGGGTCTCTGGAGTACCAAGCCGCTGTTCTCTCTGTGAGATGACAAAT 1656
QY 394 AspAsnAlaProGlnPheSerGlnGlySerArgTyrValGlnValArgGluAspValThr 413
DB 1655 GATTAAGCCCCCAGATTAGTAGAAGCCCTAATGTGTCTAGGTGAGGAGATGTGACT 1596
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433
DB 1595 CCAGGGGCCCGAGTACTCCGAGTACAGCCCTCGAATCGAGACAAAGGGAGCAATGCCGTG 1536
QY 434 ValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThr 453
DB 1535 GTGCACTATAGCATCATGAGTGGCAATGCTCGGGGACAGTTTATCTGTGATGCCCACT 1476
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrGluThrThrIleGlnTyrThrLeuArg 473
DB 1475 GGAAGCTCTGGATGTGTGAGCCCTCTTGACTATGAGACCAAGAGATACACCTTACCG 1416
QY 474 ValArgAlaGlnAspGlyGlyIleArgProProLeuSerAsnValSerGlyLeuValThrVal 493
DB 1415 GTGCCAGCACAGAGATGTGGCCGCTCCCACTCTAATGTCTGTGGCTTGTGACAGTA 1356
QY 494 GlnValLeuAspAlaLeuAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513
DB 1355 CAGGTCTCGGATATCAACCAATGCCCCCATCTTGTGAGCAACCCCTTTCAGAGCTACT 1296
QY 514 ValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAsp 533
DB 1295 GTCTCGAGAGAGTGTCCCTTAAGCTACTGTCTTCCATGTCCAGGCTTACAGCTGAT 1236
QY 534 AlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPhe 553
DB 1235 GCTGTGTGCAATGCCCGCTGGAATACCGCTTGTGGGTGGGACATGACTTCCCTTC 1176
QY 554 ThrIleAsnAsnGlyThrGlyTyrPheIleSerValAlaAlaGluLeuAspArgGluGluVal 573
DB 1175 ACCATCAACAATGGACCAAGCTGAGANTCTGTGGCTGTGAACCTGACCGGAGGAAGTT 1116
QY 574 AspPheTyrSerPheGlyValGluIleArgAspHisGlyThrProAlaLeuThrAlaSer 593
DB 1115 GATTTCTACAGCTTTGGGTTAGAGCTGAGACATGGCACTCCAGCACTCACTGCTCTG 1056
QY 594 AlaSerValSerValThrValLeuAspValLeuAspAsnAsnProThrPheThrGlnPro 613
DB 1055 GCCAGTGTACAGCTGCTGCTGATGTCAACGACAAACAACTTTCACCAACCA 996
QY 614 GlnTyrThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValValThrValSer 633
DB 995 GAGTACACAGTCCGCTCATATAGATGAGTGTGGGACCAAGGTGTGTGACGGTGTCA 936
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArg 653
DB 935 GCTGTGACCGGTGATGCTCATGTGTCACTTACCAAGATCAACAGTGGCAATACTGA 876
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673
DB 875 AACCCCTCTCCATACCAAGCAAGTGTGTGGGTGTGATCCCTTGCCTGCACTG 816
QY 674 AspTyrValLeuGluAspGlnTyrValLeuAlaValThrAlaSerAspGlyThrThrGln 693
DB 815 GACTTCAAACTTGAGGGGAGATGTGTGTGTGTACCGCTCCGATGGCACTCGGGCAG 756
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713
DB 755 GACACGGGACAAATTTGTGTGATGTCAACCGACGCGCAACCACTGCTCTGTCTTACG 696

```

```

QY 714 SerSerHisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu 733
DB 695 AGCTCCCACTATACAGTAAATTTTATGAGAGCCGGCCGGGACAGCAACAGGTGTGCTG 636
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAsp 753
DB 635 ATCAGCGCCAGAGAGAGACACAGGTGAGATGCCCACTCACTTCAATGAGAGAC 576
QY 754 SerIleProGlnPheArgIleAspAlaAspThrArgValAlaValThrThrGlnAlaGlnLeu 773
DB 575 AGCATCCCCCAGTTCGATGATGATCAACACGGGGGCTGTACACCAAGGCTGAGCTG 516
QY 774 AspTyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793
DB 515 GACTATGAAGACCAAGTGTCTTACACCTGGGCATTAATGTCTGGGACATATGCAATTCCC 456
QY 794 GlnIleSerAspThrThrTyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro 813
DB 455 CAGAGTCCGACACCACTTACCTGAGATCCTGTGTGAACGATGAAATGACATATGCCCT 396
QY 814 GlnPheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSer 833
DB 395 CAGTTCCTCGAGACTCTTACCAAGGCAGTGTCTATGAGATGTGCCACCTTACCTAGC 336
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThr 853
DB 335 GTCTCGACAGTCTCAGCCACTGATCGTATCTGGACTTAATGGCAGGTCTTTCACACC 276
QY 854 PheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleVal 873
DB 275 TTCAGAGAGGCGACGAGTGAAGAGACGTTATTTGTGTGATCCACTCAGGCAATCGTG 216
QY 874 ArgThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAla 893
DB 215 CGAACGTCAGGAGGCTGATCGAGAACCTGGCCCATATGTCTTCCGGCATATATCA 156
QY 894 ValAspIleGlyMetProProAlaArgThrProMetGluValThrValThrValLeuAsp 913
DB 155 GTGACAAAGGGAGATGCCCGCCAGCCGACACCTATGAGATGACAGTCACTGTGTGAT 96
QY 914 ValAsnAspAsnProProValPheGluGluAspGluPheAspValPheValGluGluAsn 933
DB 95 GTGAATGCAATCCCCCTGTCTTGTGACAGATGATGATTTGATGTGTGTGAAGAAGAC 36
QY 934 SerProIleGlyLeuAlaValAlaArgValThr 944
DB 35 AGCCCATTTGGGCTAGCCGTGGCCCGGCTACA 3

```

RESULT 9  
 US-10-125-540-570/c  
 ; Sequence 570, Application US/1012540  
 ; Publication No. US2003059875A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT121A1  
 ; CURRENT APPLICATION NUMBER: US/10/125,540  
 ; PRIORITY FILING DATE: 2002-04-19  
 ; Prior Application removed - See File wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 646  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 570  
 ; LENGTH: 2077  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-125-540-570

Alignment Scores:  
 Pred. No.: 0  
 Score: 691.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 23.64%

Length: 2077  
 Matches: 691  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0

DB: 15 Gaps: 0  
US-09-916-849a-3 (1-2923) x US-10-125-540-570 (1-2077)  
QY 254 ThHhIvAlPheArYvAlThrAlaGlnAspHisGlyWecProArGArSeraAlaLeuAla 273  
Db 2075 ACCCAGGCTCTTCAGGCGCAAGCGCGCAAGCAAGCGCGCTGCCCCGCAAGATGCGCTGGCT 2016  
QY 274 ThHleuThrIleuValThrAspThrAspAspHisAspProValPheGluGlnGlu 293  
Db 2015 ACCTCAACCACTCTGGTACTGACCAACCAATGACCAATGACCTGTGTGTGAGCAAGAG 1956  
QY 294 TyLysGluSerLeuArgGluAenLeuGluValGlyTyrgluValLeuThrValArgAla 313  
Db 1955 TACAAGAGAGCGCTCAGAGAGAACCTGAGGTGGCTATGAGGGTCACTGTCAAGGCC 1896  
QY 314 ThrAspGlyAspAlaProProAlaAsnIleLeuTyArgLeuLeuGluGlySerGly 333  
Db 1895 ACGGATGTGATGCCCTCCCAATGCCAAATTCATGACCGCTGCTGAGGGGTCTGGG 1836  
QY 334 GlySerProSerGluValPheGluIleAspProArGserGlyValIleArGThrArgGly 353  
Db 1835 GCGAGCCCTCTGAAGCTTGTGAGATGACCTGCTGTGGGTGATCCAAACCCGTGAC 1776  
QY 354 ProValAspArgGluGluValGluSerTyrgluLeuThrValGluAlaSerAspGlnGly 373  
Db 1775 CCGTGGATCGGGAAGAGTGGATCTTACCAAGCTGACGGTATGAGGCAAGTCAAGGGT 1716  
QY 374 ArgAspProGlyProArGserThrThrAlaAlaValPheLeuSerValGluAspAspAsn 393  
Db 1715 CGGAGCCCGGGTCTCTGAGATACCAAGCGCTGTCTTCTGTGAGGATGACAAAT 1656  
QY 394 AspAsnAlaProGlnPheSerGluLysArgTyrgluValGlnValArgGluAspValThr 413  
Db 1655 GATTAATCCCTCCAGATTAGTGAAGACCTATATGTCCAGGTGAGGAGAGTGTGACT 1596  
QY 414 ProGlyAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaVal 433  
Db 1595 CCAAGGGGCCCAAGTACTCCAGTCAAGCTCGGATCGAGCAAGGGGAGCAATGCGGTG 1536  
QY 434 ValHisTyrgSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrgLeuAspAlaGlnThr 453  
Db 1535 GTGCACATATGACATGATGTGCATGTGCGGAGACATTTATCTGATGAGCCAGACT 1476  
QY 454 GlyAlaLeuAspValValSerProLeuAspTyrgluThrThrylGlnTyrgThrLeuArg 473  
Db 1475 GAGGCTCTGATGTGTGAGACCTCTTGAATGAGACGACCAAGAGTACACCTTACGG 1416  
QY 474 ValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
Db 1415 GTCCGAGCACAAGATGTGTGCGCGTCCCACTCTCTAATGTCTCTGGCTTGGTGAAGTA 1356  
QY 494 GlyValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
Db 1355 CAGGTCTGTGATATCAACGACATGCCCCCAATCTTCTGTGACGACCCCTTTCAGGGCTACT 1296  
QY 514 ValLeuGluSerValProLeuGlyTyrgLeuValLeuHisValGlnAlaIleAspAlaAsp 533  
Db 1295 GTCTGAGAGGTGCCCTTATAGCTACTGTGTTTCCATGTCCAGGCTATCGAGCTGAT 1236  
QY 534 AlaGlyAspAsnAlaArgLeuGluTyrgLeuAlaGlyValGlyHisAspPheProPhe 553  
Db 1235 GCTGTGTACATGTCCCGCTGAGATACCGCTTGTGTGGGTGGGACATGACTTCCCTTC 1176  
QY 554 ThrIleAsnAsnGlyThrGlyTyrgIleSerValAlaIleGluLeuAspArgGluGluVal 573  
Db 1175 ACCATCAACATGCAACAGCTGATCTCTGTGCTCTGACTGACCGGAGAGAGATT 1116  
QY 574 AspPheTyrgSerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSer 593  
Db 1115 GATTTTACAGCTTTGGGGTGAAGCTCGAGACATGAGCACTCCAGACTCACTGCTCG 1056  
QY 594 AlaSerValSerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnPro 613

Db 1055 GCCAGTGCAGCGTGCATCTCTCGATGTCAACGACCAACATTCACCTTACCCACCA 996  
QY 614 GlnTyrgThrValArgLeuAsnGluAspAlaAlaValGlyThrSerValThrValSer 633  
Db 995 GAGTACACAGTGGCTCATGATGAGATGCGATGTGGGACCAAGCGTGTGACGGTGA 936  
QY 634 AlaValAspArgAspAlaHisSerValIleThrTyrglnIleThrSerGlyAsnThrArg 653  
Db 935 GCTGTGACCGTGAATGCTCATAGTGTCACTTACATGACATGACATGACATGACATGAC 876  
QY 654 AsnArgPheSerIleThrSerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeu 673  
Db 875 AACCGCTTCTTCATCACGACGCAAGTGGTGGCTGTATCCCTTCCCTCCGACACTG 816  
QY 674 AspTyrgLeuGluGluArgGlnTyrgValLeuAlaValThrAlaSerAspGlyThrArgGln 693  
Db 815 GACTCAAACTTGAAGCGGAGTATGTGTGGCTGTATCCGCTCCGATGGCACTCGGACG 756  
QY 694 AspThrAlaGlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGln 713  
Db 755 GACACGGCACAATTTGTGTGATGTCAACCGACCGCAACCAATGTCTGTCTTTCAG 696  
QY 714 SerSerHisTyrgThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeu 733  
Db 695 AGCTCCCACTATACATGATGTATATGAGACCGGCGCGGACGACCAAGGTGTGCTG 636  
QY 734 IleSerAlaThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrgPheMetGluAsp 753  
Db 635 ATCAGCGCCACGATGAGAGACACAGGTGAGATGCCCGCATCACCTCATTCATGAGGAC 576  
QY 754 SerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeu 773  
Db 575 AGCATCCCACTTCCGATTCGATGATGACAGACAGGGGGCTGTACCAACCGACTGAGCTG 516  
QY 774 AspTyrgluAspGlnValSerTyrgThrLeuAlaIleThrAlaArgAspAsnGlyIlePro 793  
Db 515 GACTATGAAAGACCAAGTGTCTTACACCTCGGCAATTAAGCTCGGAGCAATGCACTTCCC 456  
QY 794 GlnLysSerAspThrThrTyrgLeuGluIleLeuValAsnAspValAsnAspAsnAlaPro 813  
Db 455 CAGAGTCCGACACCACTTACTGAGATCTGTGTGAGACAGATGATATACATATGCCCT 396  
QY 814 GlnPheLeuArgAspSerTyrglnGlySerValTyrgluAspValProProPheThrSer 833  
Db 395 CAGTTCTGCGAGACTCTTACACAGGGAGGTCTTATGAGATGTGCAACCTTCACTTAC 336  
QY 834 ValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrgThr 853  
Db 335 GTCTGTGACATCTCAGCCCACTGATGTGATCTGTGACATTAATGCGAGGGTCTTCAACC 276  
QY 854 PheGlnGlyValAspAspGlyAspPheIleValGluSerThrSerGlyIleVal 873  
Db 275 TTCAGAGAGGCGCAGATGAGACGCTGACTTTATTTGATGATCAGCTCAGGCACTGTG 216  
QY 874 ArgThrLeuArgArgLeuAspArgGluAenValAlaGlnTyrgValLeuArgAlaTyrgAla 893  
Db 215 CGAACGCTACGAGAGGTGATCGAGAGACGTGCGCCAAATATGTCTTGGGCGCATATGCA 156  
QY 894 ValAspIleGlyWecProProAlaArgThrProMetGluValThrValThrValLeuAsp 913  
Db 155 GTGAGACAGGGAGTGGCCCCCAGCCGACACCTATGAGATGACATGACTGTGTGAGAT 96  
QY 914 ValAsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGluAen 933  
Db 95 GTGATGACAAATCCCTGTCTTGTGACAGAGATGATGATGTGTGTGTGGAAGAGAAC 36  
QY 934 SerProIleGlyLeuAlaValAlaArgValThr 944  
Db 35 AGCCCATTTGGGCTGAGCGTGGCCCGGTGCACA 3  
RESULT 10  
US-09-843-856-1

```

Sequence 1, Application US/09843856
Patent No. US20020034785A1
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R.
LI, YI
RUBEN, STEVEN M.
TITLE OF INVENTION: CALCITONIN RECEPTOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSER: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/843,856
FILING DATE: 30-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/970,758
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: STEPPER, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.066001/EKS/KMT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2603 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 652..2355
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-843-856-1
Alignment Scores:
Pred. No.: 0 Length: 2603
Score: 537.00 Matches: 723
Percent Similarity: 99.45% Conservative: 0
Best Local Similarity: 99.45% Mismatches: 2
Query Match: 18.37% Indels: 4
DB: 9 Gaps: 0
US-09-916-849a-3 (1-2923) x US-09-843-856-1 (1-2603)
QY 2197 AADPLEUGLThrThrValIleuProGluSerValPheArgGluThrProProVal 2216
DB 175 GACCTTGAGACACATCTATCTGCTGAGTCTGTCTTCAAGAGAGCGCCCGCGGTGTC 234
QY 2217 ATGProAlaGlyProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGlnArg 2236
DB 235 AAGCGCCGAGGCGCCCGAGAGGCGCCAGAGCCAGAGAGCTGGCACGCGCAGAGCGG 294
QY 2237 HAsProGluLeuSerGlnGlyGluAlaValAlaSerValIleIleTyArgThrLeuAla 2256
DB 295 CACCCGAGACTGACGAGGAGGTGAGGCTGTGGCCAGGCTCATCTACCGCACCCCTGGGC 354
QY 2257 GtLeuLeuProHisAsnTyraProAspLysArgSerLeuArgValProLysArgPro 2276
DB 355 GGGCTACTGCTCATTAATAACCTTGACAGCGGAGCTTGAGAGTCCCAAGGCGCG 414

```

```

QY 2277 IleIleAsnThrProValValSerIleSerValHisAspAspGluGluLeuLeuProArg 2296
DB 415 ATCATCAACACACCCGTGTGAGCATCAACGCTCCATGATGATGAGAGCTTGTGCCCCG 474
QY 2297 AlaleuAspLysProValThrValGlnPheArgLeuLeuGluThrGluGluArgThyLys 2316
DB 475 GCCCTGACAAACCCGTCAAGGTGAGTTCCGCTGTGAGACAGAGAGGCGACAG 534
QY 2317 ProIleCyValPheThrAsnHisSerIleLeuValSerGlyThyGlyTyrThrAla 2336
DB 535 CCCATCTGTGTTTGTGAAACATTCATCTCTGTGACAGGAGGACAGTGTGTGTGGCC 594
QY 2337 ArgGlyCyGluValValPheArgAsnGluSerHisValSerCyGluLys-AsnHisWe 2356
DB 595 AAGAGCTGTGAAGTGTCTCCGCAATAGAGACCGACGTCAGTCCAGT-CACCACT 653
QY 2356 ThrSerPheAlaValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLe 2376
DB 654 GACGAGCTTCGCTGTGCTCATGAGCGTTTCTGCGGAGAAAGGAGATCCTGCACT 713
QY 2376 WtyThrLeuThrTyraValAlaLeuGlyValThr-LeuAlaAlaLeuLeuThrPheP 2396
DB 714 GAAGACACTGACATACGTGCTAGGTGAGC-CTTGAGTCCCTTGTGCTACCTTCT 772
QY 2396 hPheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuT 2416
DB 773 TCTTCTCATCTCTTGGGTATCTCCGCTCCACCAACAGCGCATCCGACGTATCTGA 832
QY 2416 hAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuP 2436
DB 833 CAGCTGCGCTGGCGCTGCTGAGCTGAGTCTTCTCTCGGAAATCAACAGGCTGACCTCC 892
QY 2436 RobPheAlaCyThrValIleAlaIleLeuLeuHisPheLeuTyrrPheSerT 2456
DB 893 CTTTGCTGACATCAATGCAATGCAATCTGTGCACTTCTCTGACCTGTGCACTTTTCT 952
QY 2456 rPalaleuLeuGluAlaLeuHisIleuTyraAlaLeuThrGluValArgAspValAsnT 2476
DB 953 GGGCTCTGCTGAGGCTTGGACCTGTACCGGGACCTCACTAGAGTGCGCATGTCAACA 1012
QY 2476 hTgLyProMetArgPheTyrrMetLeuGlyTyrGlyValProAlaPheIleThrGlyL 2496
DB 1013 CCGGCCCATGCGCTTCACTACATGCTGGGCTGGGCGTGGCTGCTCCATCAACAGGGC 1072
QY 2496 euAlaValAlaGluAspProGluGlyTyrrGlyAsnProAspPheCyrrPheSerIleT 2516
DB 1073 TAGCGTGAGCTGAGCCCGAGGCTAGGAGAACCTTGACTTGTGTGCTGCTCATCT 1132
QY 2516 YrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSerValP 2536
DB 1133 ATGACAGCTCATCTGAGAGTTTGTGTGCGCCGAGTGGCTTTGGCTTCGATGAGTGTCT 1192
QY 2536 hLeuTyrrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGluTyL 2556
DB 1193 TCTGTATCATCTGCGCGCGCGGCTCTGTGTGCTGCCACAGCGGAGGCTTTGAGAAGA 1252
QY 2556 YrGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrT 2576
DB 1253 AAGGTCTGTCTCGGAGCTGAGCGCTCTTGTGCGGCTCTCTGTGTGTGAGCGCAGCT 1312
QY 2576 rPLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrrLeuPheAlaT 2596
DB 1313 GCGTCTGTGCACTGTCTCTGTCAACAGCGACACCTCTCTTCCACATACCTTTGTGTA 1372
QY 2596 hTyraSerCyellGlnGlyProPheIlePheLeuSerTyraValLeuSerLysGluV 2616
DB 1373 CCTGCAATGTGATCAAGGCGCCCTTCATCTTCTCTATGTGTGTGTGTGCAAGGAGG 1432
QY 2616 aArgLysAlaLeuLysLeuAlaCySerArgLysProSerProAspProAlaLeuThrT 2636
DB 1433 TCCGAAAGCACTCAAGTTCCTGTGAGCCCGAAGCCAGGCTTGTGACCTCTGTGACA 1492
QY 2636 hTyrSerThrLeuThrSerSerTyraSerTyraSerProTyraAlaAspGlyArgLeuT 2656

```

```

Db      1493 CCAAGTCCACCCGACCTCGTCTACCACTGCCAGCCCTCAAGCAAGATGGCGCTGT 1552
Qy      2656 YGGINPOTRYGIIYAASPSERIALGIYSERLEUHISSERTHRSSERARGSERGIYLVSSERG 2676
Db      1553 ACCAGCCCTTACGAGACTCGGCGGCTCTGTGACAGCAACATCGCTCGGCAAGAGTC 1612
Qy      2676 INPROSETRYIIPEOPHELEULEARGIUGIUSERIALALEUAAANPROGIYGINIYP 2696
Db      1613 AGCCCAAGCTACATCCCTTCTTGTCTAGAGGAGAGTCCGACATGAAOCCCTGGCAAGGCG 1672
Qy      2696 ROPTROGIYLEUGIYAAPPROGIYSERLEUHPHELEUGIUGIYINAPGININHIASPP 2716
Db      1673 CCCCTGGCCCTGGGCGATCCAGAGCAGCTGTCTCGGAAGGTCAAGACAGCAACATGATC 1732
Qy      2716 ROASPHTRAPSERAPSERAPSERAPSERLEUGIUNAPAPGINISERGIYSETRYIIAS 2736
Db      1733 CTACACAGGACTCGAGAGTGAAGCTGTCTTAAAGACAGACAGAGTGTCTCTATGCTT 1792
Qy      2736 ERTHRHISSESEASPSERGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGIUGI 2756
Db      1793 CTACCCACTCATCAGACAGTGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1852
Qy      2756 ROGIIYUGIUGIYTPAPPSERLEULEUGIYPROGIYALAGIUAAGLEUPROLEUHISS 2776
Db      1853 CTGGAGAGCAGGCGCTGGATAGCTGCTGGGCGCTGGAGCAGAGAGACTGCCCTGCACA 1912
Qy      2776 ERTHRPROLYSAPSGIYGIYPROGIYPROGIYLYSALAPOTRPPROGIYAAPSPHGIYT 2796
Db      1913 GATCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1972
Qy      2796 HRTHRALYSGIUSERSERGIYASNGIYALAPROGIUGIUAAGLEUARGIYUASNGIYA 2816
Db      1973 CCAAGCAAAAGAGAGTGAAGTGGACAGGCGCCCTGAGAGAGAGAGAGAGAGAGAGAGAG 2032
Qy      2816 SPALALEUSERRAAGIUGIYSERLEUGIYPROLEUPROGIYSESERIALAGINPROHIAL 2836
Db      2033 ATCCCTGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2092
Qy      2836 YSGIYIIIEULYSLYLYSEYLEUPROTHRIIESERGIYUSERSERLEULEUARGI 2856
Db      2093 AAGGCACTCTTAAGAGAGAGTGTCTGCCACCATCAGCAGAGAGAGAGAGAGAGAGAGAG 2152
Qy      2856 EUPROLEUGIUGIYSEYTHGIYSESERIARGIYSESERIALASERGIUGIYSEIARG 2876
Db      2153 TCCCTCGAGAGCAATGACAGAGGCTTCCCGGAGCTCTCCGCTAAGAGAGAGAGAGAG 2212
Qy      2876 IYGIYPROPTROTHARGPROPTROTHARGIINSEIRLEUGIUGIINLEUANGIYVALM 2896
Db      2213 GCGGCCCCCTCCCGCCGACCGCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2272
Qy      2896 ECPROLIETALAMETSERILEYVALAGIYTHRVALLAPAGIUAASPSERGIYSEIARG 2916
Db      2273 TGCCCATCGCATAGCATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2332
Qy      2916 HELEUPHEPHEANPHE 2921
Db      2333 TTCTCTTCTTAACCTTC 2349

```

```

; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 636
; LENGTH: 2695
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65)..(65)
; OTHER INFORMATION: n equals a,t,g, or c
;
; NAME/KEY: misc feature
; LOCATION: (356)..(356)
; OTHER INFORMATION: n equals a,t,g, or c
;
; NAME/KEY: misc feature
; LOCATION: (2685)..(2685)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-10-264-237-636

Alignment Scores:
Pred. No.: 1,456-266 Length: 2695
Score: 281.00 Matches: 639
Percent Similarity: 97.86% Conservative: 0
Best Local Similarity: 97.86% Mismatches: 7
Query Match: 9.61% Indels: 14
DB: 12 Gaps: 0

US-09-916-849a-3 (1-2923) x US-10-264-237-636 (1-2695)

Qy      2278 ILEANTHRPROVALVALSERILESERVALHISAPAPGIUGIULEUPROARGALA 2297
Db      485 ATCAACACACCCGCTGTGAGCATCAGCGTCCATGATGATGAGAGACTTCTGCCGCGGCC 544
Qy      2298 LEUAPPLYPROVALTHRVAGIINPHEARGLEULEUGIUTHTGIIUGIUAARGIYSPRO 2317
Db      545 CTGACAAACCCGTACAGGTGCACTTCCGCTGTGAGACAGAGAGAGAGAGAGAGAGAGAG 604
Qy      2318 ILECYVALPHEPTRAPASNHISERILEUVALSERGIYTHRGIYIYTRPSERIALARG 2337
Db      605 ATCTGTCTTCTTGACATCATTCATCTGTGATGACAGAGAGAGAGAGAGAGAGAGAGAG 664
Qy      2338 GIYSGIYVALVALPHEARGANGIUSERHISVALSERCYSGIINCYE-ASNHISMETH 2357
Db      665 GCGTGTGAAGTGTCTTCCGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 723
Qy      2357 RSERPHEALVALLEUMETASPPVALSERIARGIUGIUNASNGIYGIYIIELEUPROLEU 2377
Db      724 GAGCTTCGCTGTCTCATAGAGAGTTCCTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 783
Qy      2377 ERTHRLEUTHRYVALALALEUGIYVALTHR-LEUALALALEULEUETHRRPHEPHEP 2397
Db      784 GACATGACATACGAGGCTGTAGGTGCGC-CTTGCGTCCCTTCTGCTCACTTCTTCT 842
Qy      2397 HELEUTHRIEULEUARGIILEUARGSERANGIINHIISGIYIIEARGARGAENLEUTHRA 2417
Db      843 TCTCATCTCTTGTGGATCTCGCGCTCCAAACACAGCGCATCCGACCTTAACCTGACAG 902
Qy      2417 IALALEUGIYLEUALAGIINLEUVALPHELEULEUGIYIIEANGIINIAAPLEUPROP 2437
Db      903 CTGCGCTGGGCTGTGCTAGCTGAGTGTCTTCTCTGGAGATCAACAGGCTGACCTCCCTT 962
Qy      2437 HEALA-CYETHRVALLIETALILEULEUHISSPHELEUYRLEUCYETHRPHSEERTP 2456
Db      963 TTGS-CTGACAGTCAATGCCATCCGTGCACTTCTGTACTCTGCACTTTTCTG 1021
Qy      2457 ALALEULEUGIUALALEUHISSLEUYRARGALALEUTHGIUVALARGSPVALASNTHR 2476
Db      1022 GCTGTGTGAGGCTGTGACCTGTACCGGAGCACTACAGAGGTGCGCATGTCAACACC 1081
Qy      2477 GIYPROMETARGPHEITYRMECTLEUGIYTPGIYVALPROALAPHEIETHRGILEU 2496
Db      1082 GGCCCATCGCTTCTACTACATGCTGGGCTGGGCGTGCCTGCTTATCAAGGCGTA 1141

```

```

RESULT 11
US-10-264-237-636
; Sequence 636, Application US/10264237
; Publication No. US2004009491A1
; GENERAL INFORMATION:
; APPLICANT: Biree et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19

```

QY	2497	AlAVeAlGIvLeuAaSPProciugLiYTrGIvAaPProaPheCyETrPLeuSerLiETrY	2516
Db	1142	GCGGGGGGCTGACCCCGAGGGACTGACGGAAACCTGACCTTGTGTGGCTCTCCATCAT	1201
QY	2517	AsPTrhLeuLiETrPSePheAla-GIProVaAlaPheAlaValSerMetSeVaPh	2536
Db	1202	GACAGGCTCATCTGGAGTTTGG-TGGCCCGGTGGCTTGGCGTCTGAGTAGGTCTT	1260
QY	2536	eLeuTrYLleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGlnuYAlY	2556
Db	1261	CCGTGACATCCGGGGGGCCCGGGGGCTCTGTCTCCAGCGGCGGCTTGGAAAGA	1320
QY	2556	agLYProVaLSerGlyLeuGlnProSePheAlaValleuLeuLeuSeRaLaThTr	2576
Db	1321	AGGTCTGTCTGGGCTGGAGCTTCCTTCCCGGTCTCTGTCTGGAGGCCACTG	1380
QY	2576	pleuLeuAlaLeuLeuSeRaValaSnSer-AsPTrhLeuLeuPheLiETrYleuPheAla-	2595
Db	1381	GCTGTGGCACTGCTCTGTCAACAG-AGACACCTCTCTTCCACTACCTTTGS-T	1438
QY	2596	ThrCyAaNCyAlaEgGlnGlyProPheLiELeuSeTrYValaValleuSerLyGln	2615
Db	1439	ACCTGCAATTGATCAGAGGGCCCTTCATCTTCTCTATGTGTGCTTAGCAAGAG	1498
QY	2616	ValArgLYAlaLeuLYleuAlaCySeSerArgLYsProSeProaSProAlaELeuThr	2635
Db	1499	GTCGGAAAGCACTCAAGCTTGCTCGAGCGGCAAGCCCAAGCCCTGACCTCTGAC	1558
QY	2636	ThrLYsSerTrhLeuThSerSeTrYTrAaNCyProSeProTrYAlaAaPGLYArgLeu	2655
Db	1559	ACCAAGTCACCCCTGACCTCGTCTTCAACATCGCCCAAGCCCTTACGCAAGAGGCGCTG	1618
QY	2656	TYrGlnProTrYrGLYAaSPSeRaAGLYSeTrLeuLiHSeTrhSaArgSeRGLYSeSer	2675
Db	1619	TACAGGCCCTTAGAGACTCGGGCGGCTCTGTGACAGACCAAGTCGTCGGGCAAGGT	1678
QY	2676	GlnProSeTrYrLiEProPheLeuLeuArgGlnuSerAlaLeuAaSProGlyGlnGly	2695
Db	1679	CAGCCCAACTCATCCCTCTTGTGTGAGGAGAGTCCGACAGAACCTGGCGCAAGGG	1738
QY	2696	ProProGlyLeuGlyAaSPProGlySeTrLeuPheLeuGlnGlyGlnaPngGlnLiHAsP	2715
Db	1739	CCCCCTGGCTGGGGAGTCCAGAGAGCTGTCTCGAAGGTCAAGACAGACCAAGAT	1798
QY	2716	ProaPTrhAaSPSeRaSPSeRaPLeuSeRLeuGlnaPaaPcJnSerGlySeTrYrAla	2735
Db	1799	CCTGACACGAGCTCCGACAGTGACTGTCTTAGAAGACGACGAGTGGCTCTAAGCC	1858
QY	2736	SeTrhThHIsSeSerAaSPSeRGLnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaPhe	2755
Db	1859	TCACCCCACTATCAGACAGTGAAGAGGAAAGAGGAGGAGGAGGAGGCGGCTTC	1918
QY	2756	ProGlyGlnGlnGlyTrPAaSPSeRLeuLeuGlyProGlyAlaGlnuArgLeuProLeuHs	2775
Db	1919	CCTGAGAGGAGGCTGGGATAGCTGTGGGGCTGGAGAGAGAACTGCCCTTCAAC	1978
QY	2776	SeTrhProLYAaPGLYGlyProGlyProGlyLYaLaProTrPProGlyAaSPheGly	2795
Db	1979	AGTACTCCCAAGATGGGGGCCCAAGGGCTTGGCAAGGCCCTTGGCCAGAGACTTTGGG	2038
QY	2796	ThrThrAlaLYsGlnuSeSerGlyAaSnGlyValaProGlnGlnuArgLeuAaArgLYuAnGly	2815
Db	2039	ACCAACAGAAAGAGATGTAGCGAACCGGGCCCTTAGAGGCGGCTGGCGGAGAAATGGA	2098
QY	2816	AaPaLaLeuSeRaTrGlnGlySeRLeuGlyProLeuProGlySeSerAlaGlnProHs	2835
Db	2099	GATGCCCTGTCTCAGAGGGGTCTCCTAGAGCCCTTCAAGGCTCTTCTGCCCAAGCTCAC	2158
QY	2836	LYaGGLYleLeuLYsLYsCyLeuLeuProThrLiESeRGLYsSeSerSeRLeuAaArg	2855
Db	2159	AAAGGACATCTTAAGAAAGATGTCTGCCACCAATCAGCGAAGAGACAGCTTCTGGCGG	2218

Oy	2856	LeuProlaenGIuGInCyThrGlYSerSerArgIYSerSraIasErGIuGLYSerArg	2875
Db	2219	CTCCCCCTGGAGCANTGCACAGGGCTTCGCCGGGGCTTCCGTAGTAGAGGCGACGCCGG	2278
Oy	2876	Glycyl- ProProBProARgProProPProArgGIInserIeunGIuGLIneAenGIYa	2895
Db	2279	GGCGK- CCCCCTCCCCGCCCAACGCCGCCGCCGACAGGCTCCAGAGACAGCTGAACGGGGT	2337
Oy	2885	IMeProIIeAlametserIIelySaIagLYThValAspGIuAspserserGIysErGI	2915
Db	2338	CATGCCCATCCCACTAGGCAATGAGCGACGCGTGATGAGGACTCGTCAGGCTCCGA	2397
Oy	2915	uPhelenupheanPheanPheleuHIS	2923
Db	2398	ATTTCCTCTTTAACTTCTGTCAT	2422
RESULT 12			
US-09-925-300-359			
; Sequence 359, Application US/09925300			
; Patent No. US20020151681A1			
GENERAL INFORMATION:			
APPLICANT: Craig Rosen,			
APPLICANT: Steve Ruben			
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies			
FILE REFERENCE: PA101			
CURRENT APPLICATION NUMBER: US/09/925,300			
PRIOR APPLICATION NUMBER: PCT/US00/05988			
PRIOR FILING DATE: 2001-08-10			
PRIOR FILING DATE: 2000-03-08			
PRIOR APPLICATION NUMBER: 60/124,270			
PRIOR FILING DATE: 1999-03-12			
NUMBER OF SEQ ID NOS: 1890			
SOFTWARE: Patentin Ver. 2.0			
SEQ ID NO 359			
LENGTH: 4152			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURES:			
NAME/KEY: misc feature			
LOCATION: (65)			
OTHER INFORMATION: n equals a,t,g, or c			
US-09-925-300-359			
Alignment Scores:			
Pred. No.: 2,17e-266 Length: 4152			
Score: 281.00 Matches: 718			
Percent Similarity: 97.69% Conservative: 0			
Best Local Similarity: 97.69% Mismatches: 9			
Query Match: 9.61% Indels: 17			
DB: 10 Gaps: 0			
US-09-916-849A-3 (1-2923) x US-09-925-300-359 (1-4152)			
Oy	2197	AspleuGIuThrThrValIIleuProGIuSeValPheArgGIuThrProproValVal	2216
Db	243	GACCTTGAGACAACAGTCATCTCGCTGAGTCTGTCTTCAGAGAAGACGCCCGCGTGC	302
Oy	2217	ArgProAlaGIyProGIyGluAlaGIIngluProGIuGIuLnuLaalArgArgInaArg	2236
Db	303	AGGCGCGCACGC-CCCGAGAGGCCAGGACGACAGAGAGCTGGCA CGGCGACAGCGG	361
Oy	2237	HIsProGIuIuSerGIInglYgluaIa-ValIAserValIIlelTYrArGrThrIeuMl	2256
Db	362	CACCGGAGGTGAGCCAGGGGTAGGY-TGTGGCACGTCATCATTCACCGCACCTGGC	420
Oy	2256	aGIyLeuIeuProHIsAsnTYzAspProAbpyAsrGserIeuArgValProlysArPr	2276
Db	421	CGGGCTACTGCTCCATTAATGATGACCCTGACAGGCCAGCTTGAGAGTCCCCAAGGCC	480
Oy	2276	cIIeIIeaSnThProValValSerIIeserValHIsAspApGIuGIuLnuIeuProAr	2296
Db	481	GATCATCAACAACCCGTGTGTAGCATGACGGTCCATGATGATGAGAGACTTTCGCCCG	540

QY 2296 gAlaLeuAspLysProValThrValGlnPheArgLeuGluThrGlnGluArgThrLy 2316  
DB 541 GGCCCTGGAGAAACCCCTCAACGGGAGTTCCGCTGAGAGCAAGAGAGGGAGCA 600  
QY 2316 sProlLeuValPheThrPheHisSerIleLeuValSerGlyThrGlyGlyTrpSerAl 2336  
DB 601 GCCCATCTGTGTCTTCTGGAAACATTCAATCTGTGTAGTGGACAAGGTGGCTGGTGGC 660  
QY 2336 aaTgGlyCyAGluValPheArgAsnGluSerHisValSerCyAGlnCyS-AsnHisW 2356  
DB 661 CAGAGGCTGTGAAGTGGCTTCCGCAATGAGAGCAAGTCAAGTGGCAAGK-CAACACA 719  
QY 2356 eLTrpSerPheAlaValLeuMetAspValSerArgGluGluAsnGlyGluIleuProl 2376  
DB 720 TGAAGACCTTCGTGTCTCAATGACGTTTCTCGCGGGAATCGGGAATCTGCGCAC 779  
QY 2376 eLysPheLeuThrTrpValAlaLeuGlyValThr-LeuAlaAlaLeuLeuThrPhe 2395  
DB 780 TGAAGACCTGACATACCTGAGCTTAAAGTGTCTC-CTTGGCTGCTTCTGCTCACTTC 838  
QY 2396 PhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeu 2415  
DB 839 TTCTTCTCACTCTCTTGGATCTGCTGCTCCACCAACCAAGCAATCCGACGTAACCTG 898  
QY 2416 ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeu 2435  
DB 899 ACAGCTGCTCGGAGCTGTGCTCAGCTGGTCTCTCTCGGGAATCAACAGGCTGACCTC 958  
QY 2436 ProPheAla-CysThrValIleAlaIleLeuHisPheLeuTrpLeuCyThrPhe 2455  
DB 959 CTTTGTGS-CTGACACATCTTGGCATCTGCTGCACTTCTTACCTTCTGCACTTTTC 1017  
QY 2455 rTrpAlaLeuLeuGluAlaLeuHisLeuTrpArgAlaLeuThrGluValArgAspValAs 2475  
DB 1018 CTGGGCTCTCTGAGAGCTTGCACCTGTACCGGGACCTACGAGGTGGCATGTCAA 1077  
QY 2475 nTrpArgLysProMetArgPheTrpTrpMetLeuGlyTrpGlyValProAlaPheIleThrG 2495  
DB 1078 CACCGGCTCATGGCTTCTACTACATGCTGGGCTGGGCTGGCTCTTCTTCAACAGG 1137  
QY 2495 yLeuAlaValAlaGlyLeuAspProGluGlyTrpGlyAsnProAspPheCyTrpLeuSerI 2515  
DB 1138 GCTAAGCTGTGGCTGAGACCCCAAGGCTACGGGAACCTGACTTGTGCTGCTTCCAT 1197  
QY 2515 eTrpAspThrLeuIleTrpSerPheAla-GlyProValAlaPheAlaValSerMetSery 2535  
DB 1198 CTATGACACGCTCATCTGAGATTTCG-TGGCCGCTGCTTGTGCGTCTCATGAGT 1256  
QY 2535 aLLeuLeuTrpIleLeuAlaAlaArgLaserCyAlaAlaGlnArgGlnGlyPheGlu 2555  
DB 1257 TCTTCTGTACATCTGGCGGCTTCCGCTGTGTGCTGCCAGCGGCAAGGCTTTGAGA 1316  
QY 2555 yAlaGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT 2575  
DB 1317 AGAAGGCTCTGTCTGGGCTGACGCTCTTCCGCTGCTCTGCTGCTGAGGCCA 1376  
QY 2575 hTrpLeuLeuAlaLeuLeuSerValAsnSer-AspThrLeuLeuPheHisTrpLeuPhe 2594  
DB 1377 CGTGGCTGCTGAGCTGCTCTGTCAACAG-AGACACCTCTCTTCCACTACTCTTT 1435  
QY 2595 Ala-ThrCysAsnCyTrpIleGlnGlyProPheIlePheLeuSerTrpValValLeuSerLy 2614  
DB 1436 GS-TACCTGCAATTCATCCAGGCCCCCTTCAATCTCTCTCTAGTGTGTAGCAA 1494  
QY 2614 sGluValArgLysAlaLeuLysValaCySerArgLysProSerProAspProAlaLe 2634  
DB 1495 GAGAGCTCGGAAAGCACTCAAGGCTGCTCAAGCCGGAAGCCGAGCTGCTGTCT 1554  
QY 2634 uTrpTrpLysSerThrLeuThrSerSerTrpAsnCyProSerProGlyAlaAspGlyTr 2654  
DB 1555 GACCAACAACTCAACCTGACCTGTCTCAACTGCCCCCAAGCCCTTACCAATGGGCG 1614  
QY 2654 gLeuTrpGlnProGlyLysPheSerAlaGlySerLeuHisSerThrSerArgSerGly 2674

DB 1615 GCTGTACCAAGCCCTTACGAGACTCGGCTCTCTGCAACAGCACTGCTCGGCA 1674  
QY 2674 sSerGlnProSerTrpIleProPheLeuLeuArgGluSerAlaLeuAsnProGly 2694  
DB 1675 GAGTCAAGCCCACTACATCCCTTCTTCTGAGGAGAGTCCGCACTAAACCTTGGCA 1734  
QY 2694 nGlyProProGlyLeuGlyAspProGlySerLeuPheLeuGluGlyGlnAspGlnHis 2714  
DB 1735 AGGAGCCCTGAGCTGGGAGATCCAGGCAAGCTGTCTTCTGAAAGTCAAGACAGCA 1794  
QY 2714 sAspProAspThrAspSerAspSerAspLeuSerLeuGluuAspAspGlnSerLy 2734  
DB 1795 TGATCTGACAGAGCTCCGACATGACCTGTCTTAGAAGACGACCAAGTGGCTCTTA 1854  
QY 2734 rAlaSerThrHisSerSerAspSerGluGluGluGluGluGluGluGluGluGlu 2754  
DB 1855 TGCCTTACCACTCATCAACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914  
QY 2754 aPheProGlyGluGluGlyTrpAspSerLeuLeuGlyProGlyAlaGluArgLeuProl 2774  
DB 1915 CTTCCCTGAGAGAGAGGCTGGATAGCTGCTGGGCTGTGAGCAGAGAGAGAGAGAG 1974  
QY 2774 uHisSerThrProLysAspGlyGlyProGlyLysValaProTrpProGlyAspPh 2794  
DB 1975 GACAGTACTCCCAAGAGATGGGAGGCTGAGGCTGAGAGAGAGAGAGAGAGAGAG 2034  
QY 2794 eGlyThrAlaLysGlyLysSerGlyAsnGlyAlaProGluGluArgLeuArgLys 2814  
DB 2035 TGGAGACAGCAAGAAAGAGAGTGTGCAACGGGCTCTGAGAGAGAGAGAGAGAG 2094  
QY 2814 nGlyAspAlaLeuSerArgGlyLysSerGlyLysLeuProLeuProGlySerSerAlaGln 2834  
DB 2095 TGAAGATGCTGTCTGAGAGAGGCTCTTACGAGGCTCTTCTGAGGCTCTTCTGAG 2154  
QY 2834 oHisLysGlyIleLeuLysValaCysLeuProThrIleSerGlyLysSerSerLeu 2854  
DB 2155 TCACAAGGCACTTAAAGAAAGTGTGCTCCACCACTCAAGAGAGAGAGAGAGAG 2214  
QY 2854 uArgLeuProLeuGluGlnGlySerThrGlySerSerArgLysSerSerAlaSerGlySe 2874  
DB 2215 GCGGCTCCCTCGAGAGATGACAGAGGCTTCCGAGGCTCTCTCCGTAAGTGAAGG 2274  
QY 2874 rArgGlyGly-ProProProArgProProProArgLysLeuGlnGlnLeuAsn 2894  
DB 2275 CCGGGGCGK-CCCTCTCCCTCCCAAGCGCTCCGAGAGCTTCAAGAGCTGAGAG 2333  
QY 2894 yValMetProIleAlaMetSerIleLysAlaGlyThrValaAspGluAspSerSerGlyS 2914  
DB 2334 GGGTCAAGCCCATCGCATGACATCAAGGAGGACGGTGGATGAGACTGTCAAGGCT 2393  
QY 2914 eGluPheLeuPheAspPheAsnLeuHis 2923  
DB 2394 CCGAATTTCTTCTTTAACTTCTGCAAT 2422

RESULT 13  
US-09-764-870-571  
Sequence 571, Application US/09764870  
Patent No. US20020042386A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT214  
CURRENT APPLICATION NUMBER: US/09/764, 870  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 646  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 571  
LENGTH: 652  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-870-571



## Alignment Scores:

Pred. No.: 5,93e-204 Length: 652  
 Score: 217.00 Matches: 217  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.42% Indels: 0  
 DB: 9 Gaps: 0

US-09-916-849A-3 (1-2923) x US-09-764-870-571 (1-652)

```

QY 567 G|U|E|U|A|S|P|A|S|G|U|G|U|V|A|L|A|S|P|H|E|Y|S|E|P|H|E|G|I|V|A|G|U|A|A|A|S|P|H|I|S|G|I| 586
    |||||
DB 1 G|A|C|T|G|A|C|C|G|G|A|G|G|A|G|T|G|A|T|T|C|T|C|A|G|C|T|T|G|G|G|T|A|G|A|G|C|T|G|A|G|C|A|T|G|C| 60
    |||||
QY 587 T|H|P|R|O|A|L|E|U|T|H|A|S|E|A|S|E|V|A|S|E|V|A|T|H|V|A|L|E|U|A|S|P|V|A|A|S|P|A|S|N| 606
    |||||
DB 61 A|C|T|C|A|G|C|A|C|T|C|G|C|C|G|C|A|G|T|G|A|G|C|G|T|C|C|T|G|A|T|G|T|C|A|C|A|G|C|A|C| 120
    |||||
QY 607 A|S|P|P|H|E|T|H|P|H|G|I|N|P|R|O|G|I|U|T|Y|T|H|V|A|A|S|P|A|S|N|G|I|U|A|S|P|A|A|A|V|A|G|I| 626
    |||||
DB 121 A|A|T|C|C|A|C|C|T|T|A|C|C|A|C|A|G|A|G|T|A|C|A|G|T|C|G|C|T|C|A|A|T|G|A|G|A|T|C|A|C|T|G|G|C| 180
    |||||
QY 627 T|H|S|E|V|A|I|T|H|V|A|S|E|A|A|V|A|S|P|A|S|P|A|A|H|I|S|E|V|A|I|I|E|T|H|T|Y|G|I|N| 646
    |||||
DB 181 A|C|C|A|G|C|G|G|T|G|A|C|G|G|T|C|A|G|T|G|G|A|C|C|G|T|G|A|T|G|C|A|T|A|G|T|C|A|C|T|A|C|C|A|G| 240
    |||||
QY 647 I|L|E|T|H|S|E|G|I|A|S|H|T|H|A|S|H|A|S|P|H|E|S|E|I|I|E|T|H|S|E|G|I|N|S|E|G|I|Y|G|I|Y|L|E|U| 666
    |||||
DB 241 A|T|C|C|A|G|T|G|G|C|A|A|T|C|T|G|A|A|C|C|G|T|T|C|C|A|T|C|A|C|A|G|C|C|A|A|G|G|T|G|G|G|C|T|G| 300
    |||||
QY 667 V|A|S|E|I|U|A|I|A|E|U|P|R|O|L|E|U|A|S|P|T|Y|L|S|E|U|G|I|U|A|S|G|I|N|T|Y|V|A|L|E|U|A|A|V|A|I|T|H| 686
    |||||
DB 301 G|T|A|T|C|C|T|G|C|C|T|G|C|A|C|T|G|A|C|T|A|C|A|A|C|T|T|G|A|G|C|G|C|A|A|T|G|T|G|G|C|T|G|T|A|C|C| 360
    |||||
QY 687 A|A|S|E|A|S|P|G|I|T|H|A|S|G|I|N|A|S|P|T|H|A|I|A|G|I|I|E|V|A|I|A|S|N|V|A|I|T|H|A|S|P|A|A|S|N| 706
    |||||
DB 361 G|C|T|C|C|A|G|T|G|C|A|C|T|G|C|G|C|A|G|C|G|C|A|G|A|T|G|T|G|T|G|A|A|V|T|C|A|C|C|A|C|C|C|A|C| 420
    |||||
QY 707 T|H|H|A|S|P|P|R|O|V|A|I|P|H|E|G|I|N|S|E|S|E|H|I|S|T|Y|T|H|V|A|I|A|S|N|V|A|I|A|S|G|I|U|A|S|P|A|S|P|P|R|O| 726
    |||||
DB 421 A|C|C|A|T|C|G|T|C|C|T|G|T|T|C|T|T|A|G|A|G|C|T|C|C|A|C|T|A|C|A|G|T|A|A|G|T|A|A|G|A|G|A|C|C|G|C|G| 480
    |||||
QY 727 A|A|G|I|T|H|T|H|V|A|I|V|A|L|E|U|I|E|S|E|A|I|A|H|A|S|P|G|I|U|A|S|P|T|H|G|I|Y|G|I|U|A|S|N|A|A|A|S|G| 746
    |||||
DB 481 G|A|G|G|C|A|C|C|A|G|G|T|G|T|G|T|G|A|T|C|A|G|C|C|C|A|G|G|C|A|G|A|T|G|A|G|A|C|A|C|G|T|G|A|A|T|C|C|C|G|C| 540
    |||||
QY 747 I|L|E|T|H|T|Y|P|H|E|T|G|I|U|A|S|P|S|E|I|L|E|P|R|O|G|I|N|P|H|E|A|S|G|I|L|E|A|S|P|A|A|S|P|T|H|G|I|Y|A| 766
    |||||
DB 541 A|T|C|A|C|T|T|A|C|T|T|G|A|G|A|G|A|C|A|G|A|T|C|C|C|C|A|G|T|T|C|C|G|A|T|C|G|A|T|G|C|A|G|A|C|A|C|G|G|G|G|C|T| 600
    |||||
QY 767 V|A|I|T|H|T|H|G|I|N|A|I|G|I|U|E|U|A|S|P|T|Y|G|I|U|A|S|P|G|I|N|V|A|I|S|E|T|Y|T|H|T|H|E|U| 783
    |||||
DB 601 G|T|C|A|C|C|A|C|C|C|A|G|G|C|T|G|A|G|C|T|A|T|G|A|A|G|C|A|A|G|T|G|T|T|A|C|A|C|C|C|T|G| 651
    |||||

```

## RESULT 14

US-10-125-540-571  
 ; Sequence 571, Application US/10125540  
 ; Publication No. US20030059875A1

## GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT214C1  
 CURRENT APPLICATION NUMBER: US/10/125,540  
 CURRENT FILING DATE: 2002-04-19  
 Prior application removed - See File Wrapper or Palm  
 NUMBER OF SEQ ID NOS: 646  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 571  
 LENGTH: 652  
 TYPE: DNA

ORGANISM: Homo sapiens  
 US-10-125-540-571

## Alignment Scores:

Pred. No.: 5,93e-204 Length: 652  
 Score: 217.00 Matches: 217  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.42% Indels: 0  
 DB: 15 Gaps: 0

US-09-916-849A-3 (1-2923) x US-10-125-540-571 (1-652)

```

QY 567 G|U|E|U|A|S|P|A|S|G|U|G|U|V|A|L|A|S|P|H|E|Y|S|E|P|H|E|G|I|V|A|G|U|A|A|A|S|P|H|I|S|G|I| 586
    |||||
DB 1 G|A|C|T|G|A|C|C|G|G|A|G|G|A|G|T|G|A|T|T|C|T|C|A|G|C|T|T|G|G|G|T|A|G|A|G|C|T|G|A|G|C|A|T|G|C| 60
    |||||
QY 587 T|H|P|R|O|A|L|E|U|T|H|A|S|E|A|S|E|V|A|S|E|V|A|T|H|V|A|L|E|U|A|S|P|V|A|A|S|P|A|S|N| 606
    |||||
DB 61 A|C|T|C|A|G|C|A|C|T|C|G|C|C|G|C|A|G|T|G|A|G|C|G|T|C|C|T|G|A|T|G|T|C|A|C|A|G|C|A|C| 120
    |||||
QY 607 A|S|P|P|H|E|T|H|P|H|G|I|N|P|R|O|G|I|U|T|Y|T|H|V|A|A|S|P|A|S|N|G|I|U|A|S|P|A|A|A|V|A|G|I| 626
    |||||
DB 121 A|A|T|C|C|A|C|C|T|T|A|C|C|A|C|A|G|A|G|T|A|C|A|G|T|C|G|C|T|C|A|A|T|G|A|G|A|T|C|A|C|T|G|G|C| 180
    |||||
QY 627 T|H|S|E|V|A|I|T|H|V|A|S|E|A|A|V|A|S|P|A|S|P|A|A|H|I|S|E|V|A|I|I|E|T|H|T|Y|G|I|N| 646
    |||||
DB 181 A|C|C|A|G|C|G|G|T|G|A|C|G|G|T|C|A|G|T|G|G|A|C|C|G|T|G|A|T|G|C|A|T|A|G|T|C|A|C|T|A|C|C|A|G| 240
    |||||
QY 647 I|L|E|T|H|S|E|G|I|A|S|H|T|H|A|S|H|A|S|P|H|E|S|E|I|I|E|T|H|S|E|G|I|N|S|E|G|I|Y|G|I|Y|L|E|U| 666
    |||||
DB 241 A|T|C|C|A|G|T|G|G|C|A|A|T|C|T|G|A|A|C|C|G|T|T|C|C|A|T|C|A|C|A|G|C|C|A|A|G|G|T|G|G|G|C|T|G| 300
    |||||
QY 667 V|A|S|E|I|U|A|I|A|E|U|P|R|O|L|E|U|A|S|P|T|Y|L|S|E|U|G|I|U|A|S|G|I|N|T|Y|V|A|L|E|U|A|A|V|A|I|T|H| 686
    |||||
DB 301 G|T|A|T|C|C|T|G|C|C|T|G|C|A|C|T|G|A|C|T|A|C|A|A|C|T|T|G|A|G|C|G|C|A|A|T|G|T|G|G|C|T|G|T|A|C|C| 360
    |||||
QY 687 A|A|S|E|A|S|P|G|I|T|H|A|S|G|I|N|A|S|P|T|H|A|I|A|G|I|I|E|V|A|I|A|S|N|V|A|I|T|H|A|S|P|A|A|S|N| 706
    |||||
DB 361 G|C|T|C|C|A|G|T|G|C|A|C|T|G|C|G|C|A|G|C|G|C|A|G|A|T|G|T|G|T|G|A|A|V|T|C|A|C|C|A|C|C|C|A|C| 420
    |||||
QY 707 T|H|H|A|S|P|P|R|O|V|A|I|P|H|E|G|I|N|S|E|S|E|H|I|S|T|Y|T|H|V|A|I|A|S|N|V|A|I|A|S|G|I|U|A|S|P|A|S|P|P|R|O| 726
    |||||
DB 421 A|C|C|A|T|C|G|T|C|C|T|G|T|T|C|T|T|A|G|A|G|C|T|C|C|A|C|T|A|C|A|G|A|T|G|T|A|A|G|A|G|A|C|C|G|C|G| 480
    |||||
QY 727 A|A|G|I|T|H|T|H|V|A|I|V|A|L|E|U|I|E|S|E|A|I|A|H|A|S|P|G|I|U|A|S|P|T|H|G|I|Y|G|I|U|A|S|N|A|A|A|S|G| 746
    |||||
DB 481 G|A|G|G|C|A|C|C|A|G|G|T|G|T|G|T|G|A|T|C|A|G|C|C|C|A|G|G|C|A|G|A|T|G|A|G|A|C|A|C|G|T|G|A|A|T|C|C|C|G|C| 540
    |||||
QY 747 I|L|E|T|H|T|Y|P|H|E|T|G|I|U|A|S|P|S|E|I|L|E|P|R|O|G|I|N|P|H|E|A|S|G|I|L|E|A|S|P|A|A|S|P|T|H|G|I|Y|A| 766
    |||||
DB 541 A|T|C|A|C|T|T|A|C|T|T|G|A|G|A|G|A|C|A|G|A|T|C|C|C|C|A|G|T|T|C|C|G|A|T|C|G|A|T|G|C|A|G|A|C|A|C|G|G|G|G|C|T| 600
    |||||
QY 767 V|A|I|T|H|T|H|G|I|N|A|I|G|I|U|E|U|A|S|P|T|Y|G|I|U|A|S|P|G|I|N|V|A|I|S|E|T|Y|T|H|T|H|E|U| 783
    |||||
DB 601 G|T|C|A|C|C|A|C|C|C|A|G|G|C|T|G|A|G|C|T|A|T|G|A|A|G|C|A|A|G|T|G|T|T|A|C|A|C|C|C|T|G| 651
    |||||

```

## RESULT 15

US-09-764-870-568  
 ; Sequence 568, Application US/09764870  
 ; Patent No. US20020042386A1

## GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT214  
 CURRENT APPLICATION NUMBER: US/09/764,870  
 CURRENT FILING DATE: 2001-01-17  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 646  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 568  
 LENGTH: 509  
 TYPE: DNA

ORGANISM: Homo sapiens  
 US-09-764-870-568

## Alignment Scores:

Pred. No.:	3,87e-119	Length:	509
Score:	131.00	Matches:	168
Percent Similarity:	99.41%	Conservative:	0
Best Local Similarity:	99.41%	Mismatches:	1
Query Match:	4.48%	Indels:	1
DB:	9	Gaps:	0

US-09-916-849a-3 (1-2923) X US-09-764-870-568 (1-509)

```

QY      389 ValGluAspAspAsnAspAsnAlaProGlnPheSerGluLysArgTyrValValGlnVal 408
      |||
Db      2  GTGAGAGATGACATGATATGCCCCCAGTTTAGTAGAGAGCGCTATGCGTCCAGGTG 61
      |||

QY      409 ArgGluAspValThrProGlyAlaProValLeuArgValThrAlaSerAspArgAspLys 428
      |||
Db      62 AGGAGAGATGTGCTCCAGGGCCCCAGTACTCCAGTCCAGCTCGAGTCGAGACAG 121
      |||

QY      429 GlySerAsnAlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyr 448
      |||
Db      122 GGGAGCAATGCCGTGGTGCACTATAGCATCATGATGGCAATGCTCGGGGACGTTTAT 181
      |||

QY      449 LeuAspAlaGlnThrGlyAlaLeuAspValValSerProLeuAspTyrGluThrThrLys 468
      |||
Db      182 CTGGATGCCAGACTGGAGCTCTGATGTGTGAGGCCCTCTTGACTATGAGACGACCAAG 241
      |||

QY      469 GluTyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSer 488
      |||
Db      242 GAGTACACCTTACGGGTGCGAGACACAGATGTGGCCGCTCCCACTCTTAAATGTCT 301
      |||

QY      489 GlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThr 508
      |||
Db      302 GGCTGTGTGACAGTACAGTCTGATATCAACGACATGCCCATCTTCGTCAAGACC 361
      |||

QY      509 ProPheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuValLeuHisValGln 528
      |||
Db      362 CTTTCCAGGCTACTGCTGAGAGATGT-CCCTTAGGCTAAGCTGTTCTCCATGTCCAG 420
      |||

QY      529 AlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyrArgLeuAlaGlyValGly 548
      |||
Db      421 GCTATCGACGCTGATGCTGATGACAAATGCCCGCTGGAATACCGCTTGTGGGGTGGGA 480
      |||

QY      549 HisAspPheProPheThrIleAsnAsn 557
      |||
Db      481 CATGACTTCCCTTCAACATCAACAAT 507
      |||

```

Search completed: February 15, 2004, 03:04:58  
 Job time : 2287 sec8

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 10:45:04 ; Search time 11526 Seconds  
(without alignments)  
6163.629 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 2923  
Sequence: 1 MMSPTGVPLPPLPPLLL.....ACTVNDSSGSEFLFPFLH 2923

Scoring table: Oligo  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DBV=xlh  
-O=/cgn2\_1/USPTO.spool/US09916849/runat\_11022004\_143917\_21362/app\_query.fasta\_1.3079  
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MIMATCH=0.1 -LDOPLC=0 -LDOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849 @CGN 1 1 9354 @runat\_11022004\_143917\_21362 -NCFU=6 -ICPU=3  
-NO.WMAP -LARGEBUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-BEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELDP=6 -DELEX=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estlov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: BX409742  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_pro:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	9.4	907	13 BX409742	BX409742 BX409742
2	252	8.6	896	13 BUI73510	BUI73510 AGENCOURT
3	235	8.0	1076	12 BM480097	BM480097 AGENCOURT
4	226	7.7	908	13 BUB57213	BUB57213 AGENCOURT
5	210	7.2	695	9 AUI16945	AUI16945 AGENCOURT
6	207	7.1	662	9 ALU42580	ALU42580 DKF2P434K
7	204	7.0	877	13 BX409741	BX409741 BX409741
8	199	6.8	906	13 BQ677400	BQ677400 AGENCOURT
9	191	6.5	927	13 BQ679970	BQ679970 AGENCOURT
10	188	6.4	930	13 BQ676596	BQ676596 AGENCOURT
11	187	6.4	893	13 BUB49091	BUB49091 AGENCOURT
12	179	6.1	640	9 AW370778	AW370778 QV0-BT026
13	169	5.8	556	10 BE008566	BE008566 RCS-BN015
14	166	5.7	523	13 BX110290	BX110290 BX110290
15	165	5.6	549	10 BP951925	BP951925 QV1-NN022
16	158	5.4	477	9 AL120631	AL120631 DKF2P751I
17	151	5.2	465	10 BE012097	BE012097 RCS-BN105
18	151	5.2	498	9 AL704729	AL704729 DKF2P686H
19	151	5.2	498	13 BUI701387	BUI701387 UI-M-EXO-
20	148	5.1	647	13 BUI703750	BUI703750 UI-M-FOO-
21	142	4.9	990	13 BX460356	BX460356 BX460356
22	141	4.8	483	10 BE009645	BE009645 PM4-BN017
23	139	4.8	446	12 BI032886	BI032886 MR4-NN020
24	127	4.3	597	14 CA891550	CA891550 B016602
25	127	4.3	1018	13 BQ430341	BQ430341 AGENCOURT
26	124	4.2	405	10 BP958428	BP958428 RC4-NN117
27	122	4.2	425	10 BF092095	BF092095 RCO-TN007
28	121	4.1	616	9 AW748496	AW748496 QV0-BT026
29	115	3.9	500	2 HSM085357	AW48496 QV0-BT026
30	115	3.9	627	12 BI456703	BI456703 B03172181
31	115	3.9	722	12 BI456543	BI456543 B03172089
32	114	3.9	738	14 CA510833	CA510833 UI-R-PJ0-
33	112	3.8	812	12 BG915872	BG915872 B02815323
34	111	3.8	334	14 CB122348	CB122348 K-EST0170
35	109	3.7	480	2 HSM073596	BX483428 Homo sapi
36	107	3.7	557	10 BF369920	BF369920 UI-R-BS2-
37	106	3.6	455	10 AW900873	AW900873 CML-NN100
38	104	3.6	918	14 CA454005	CA454005 AGENCOURT
39	102	3.5	598	9 AL704278	AL704278 DKF2P686O
40	100	3.4	310	10 BE703584	BE703584 MR2-NN111
41	100	3.4	506	14 CB783412	CB783412 AMGNIC.N
42	100	3.4	694	14 CD350873	CD350873 UI-M-GTO-
43	100	3.4	846	14 CA751074	CA751074 UI-M-FOO-
44	98	3.4	621	10 AW955385	AW955385 EST367455
45	97	3.3	466	10 BF915504	BF915504 MR3-UT012

#### ALIGNMENTS

RESULT 1  
LOCUS BX409742 907 bp mRNA EST 13-MAY-2003  
DEFINITION BX409742 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
ACCESSION CSDDP013YE18 5-PRIME, mRNA sequence.  
VERSION BX409742.1 GI:30653001  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 907)

AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 131 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 64.f for more  
 information about this cluster see http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BAF012ZG07\_AF01112\_2&cluster=64.f.  
 Contact: Feng Liang Email: fliang@life.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0BAF012ZG07\_AF01112\_2.  
 Location/Qualifiers

## FEATURES

source  
 1. 907  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF013YE18"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_idb="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-Oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 BASE COUNT 148 a 322 c 238 g 198 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 5.49e-236 Length: 907  
 Score: 275.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 9.41% Indels: 0  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BK409742 (1-907)

QY 2433 AAlaAspleuProPhelaiaCyGthValiIleAlaleuLeuHsiPhelauryTleuCy9 2452  
 Db 7 GCTGACCTCCCTTGGCTGCAAGTCATTCCTGCTGCACTTCCCTGACCTCTGC 66  
 QY 2453 ThPheserTpaAlaleuLeuGluAlaleuHsiLeuTyrAraAlaLeuThrgiValArg 2472  
 Db 67 ACCCTTCTCGGGCTCTGCTGAGAGCCTTGACCTGACCGGGCACTCACTGAGGTCGC 126  
 QY 2473 AspValaAnthrgiYProMetArzPheryTtYmetLeuGlyTtTgYValProAlaPhe 2492  
 Db 127 GATGTCAACACCGGCCCATGCGCTTCTACTACATGCTGGCGGCGGCGTGCCTTC 186  
 QY 2493 IllethrgiYleuAlaValGlyleuAspProGluGlyTtYtYValAsnPProaPheCy9TTP 2512  
 Db 187 ATCAACAGGGCTAGCGCGGCTGAGACCCGAGGGCTACCGGAACCTCTGCTGCG 246  
 QY 2513 leuSerIleTyrAapThrlleuIleTtPserPheAlaGlyProValAlaPheAlaValSer 2532  
 Db 247 CTCTCCATCTATGACAGCTCATCTGAGATTTCCTGCGCCCGTGGCCTTTCGCGTCTG 306  
 QY 2533 MetSerValPheLeuTyrIleleuAlaAlaArgAlaSerCy9AlaAlaGlnATgGlnGly 2552  
 Db 307 ATGAGTCTTCTTCGTCATCTCTGCGCGGCCGCGGCTCTGCTGCTGCCAGCGAGGGC 366  
 QY 2553 PheGluYleuYserGlyProValSerGlyleuGlnProSerPheAlaValleuLeuLeu 2572  
 Db 367 TTGAGAGAGAAAGGTCTGCTCGGCGCTGAGACCTCTTCGCGCTCTGCTGCTG 426  
 QY 2573 SerAlaThrtPleuLeuAlaLeuLeuSerValaAmSerAapThrlleuLeuPheHsiTyr 2592  
 Db 427 AGGCGCACGAGTGGCTGCTGCGACCTGCTCTGTGACAGGCAACCTCTCTTCGACTAC 486

QY 2593 leuPheAlaThrCyAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValleu 2612  
 Db 487 CTCTTGTGACTGCAATTCATTCAGGGCCCTTCATCTCTCTCCATGTGATGCT 546  
 QY 2613 SerIySGlyValArgYleuAlaleuYleuAlaCy9SerArgYleuProSerProaPhePro 2632  
 Db 547 ACCAAGAGGTCCGGAAGAGCACTCAAGCTGCTGCTGACCGGCAACCCGCTCAACCT 606  
 QY 2633 AlalaThrtThySerThrlleuHsiSerTyrAraCy9ProSerProTyrAlaAap 2652  
 Db 607 GCTTGACCAACAGTCCACCTGACCTGCTCTCAACTGCGCCAGCCCTTACGAGAT 666  
 QY 2653 GlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerIleuHsiSerThsSerArgSer 2672  
 Db 667 GGGCGGCTGTACCAAGCTTACAGGAGACTCGCGGCTCTCTGACAGACAGCAGTGGCTG 726  
 QY 2673 GlyIySerGlnProSerTyrIleProPheLeuAlaGluSerAlaLeuAapPro 2692  
 Db 727 GGCAGAGTCAAGCCAGCAGTACATCCCTCTTCTGAGAGGAGAGATCCGCACTGAACCT 786  
 QY 2693 GlyIyGlnGlyProGlyleuGlyAapProGlySerIleuPheLeu 2707  
 Db 787 GGCAGAGGCCCCCTGCGCTGGGGAGATCCAGGCACTGTTCTG 831

RESULT 2  
 BUI73510 896 bp mRNA linear EST 04-SRP-2002  
 LOCUS AGENCOURT\_7939348 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6172948  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BUI73510  
 VERSION BUI73510.1 GI:22687494  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@db-rc@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM13543 row: 1 column: 05  
 High quality sequence stop: 677.

## FEATURES

source  
 1. 896  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6172948"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_idb="NIH\_MGC\_67"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Library constructed by Life  
 Technologies."

BASE COUNT 169 a 282 c 267 g 177 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.51e-215 Length: 896  
 Score: 252.00 Matches: 252  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8.62% Indels: 0

DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BU173510 (1-896)

QY 1351 PheLYSAspCYsPProSerGlyAAspPheGlyLysPProLYrCYsGluValThrThrArg 1370

DB 1 TTCAAGTGGATGGCCATCTGGAGACTTCGAGAGAGCCCTACTGCGACAGGTGACACAGCGCC 60

QY 1371 SerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThr 1390

DB 61 AGCTTCCCCGCCCACTCTTCATCACTTTGGCGGCGCGCAAGCTTCCACTTCACTTCAAC 120

QY 1331 LeuAlaLeuSerPheAlaThrIleGlyLysGluArgAspGlyLeuLeuLeuLYrAsnGlyValPhe 1410

DB 121 CTGGCCCTCTCGTTTGGCCACAAAGAGCGCGAGCGGCTGCTGTGTACAAATGGCGGTTTC 180

QY 1411 AsnGluLYrHisAspPheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPhe 1430

DB 181 AATGGAAGCATGATCTTTGTGGCCCTCGAGGTGATCCAGAGCAGGCTCCAGCTCACTTC 240

QY 1431 SerAlaGlyLysSerThrThrThrValSerProPheValProGlyValSerAspGly 1450

DB 241 TCTGCAAGGGAGGTCAACACCAACGAGTGTCCCATTCGTGCCGAGAGTCAGTATGGC 300

QY 1451 GlnTPHISThrValGlnLeuLYrTYrAsnLYrProLeuLeuGlyGlnThrGlyLeu 1470

DB 301 CAGTGGCATACCGTCCAGTCAATAATACATAAAGCACTGTGGTCCAGACAGGGCTC 360

QY 1471 ProGlnLYrProSerGlyLysGlnLYrValAlaValThrValAspGlyCYsAspThrGly 1490

DB 361 CCACAGGCGCCATTCAGAGACAGAGGTGGCTGTGGAGACCGTGGAGGCTGTACACAGAA 420

QY 1491 ValAlaLeuArgPheGlySerValLeuGlyAsnLYrSerCYsAlaAlaGlnGlyThrGln 1510

DB 421 GTGGCCTTGGCGTTGGATCTGTCTCGGGCAACTCTCTGTGCTGCCAGGGCAACCCAG 480

QY 1511 GlyGlySerLYrLYrSerLeuAspLeuThrGlyProLeuLeuGlyGlyValProAsp 1530

DB 481 GGTGCACACAAAGAACTCTCTGATCTGACGGGCGCCCTGCTACTAGCGGGGTGTCTAC 540

QY 1531 LeuProGlySerPheProValArgMetArgGlnPheValGlyCYsMetArgAsnLeuGln 1550

DB 541 CTGCCCGAGAGCTTCCCAATGCCAATGCCAGTTCGTGGGCTGCATGGGAACTTGAG 600

QY 1551 ValAspSerArgHisIleAspMetAlaAspPheIleAlaAsnGlnLYrThrValProGly 1570

DB 601 GTGCACACCGGCAATAGACATGGCTGACTTCATTGCCAACAATGGCACCGTGCCTGGC 660

QY 1571 CysProAlaLYrLYrAsnValCYsAspSerAsnThrCYsHisAsnGlyLYrThrCYsVal 1590

DB 661 TGCCCTGCCAAGAAACAGTGTGTGACAGCAACCTTGCACAAATGGGGGCACTTGGCGTG 720

QY 1591 AsnGlnTPRAspAlaPheSerCYsGluCYsProLeu 1602

DB 721 AACCACTGGAGCGCTTCACTGCCAGTGGCCCCCTG 756

RESULT 3

LOCUS BM480097 1076 bp mRNA linear EST 05-FEB-2002

DEFINITION AGENCOURT\_6468579 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5574047

ACCESSION BM480097

VERSION BM480097.1 GI:18529139

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1076)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Plate: L1AM12322 row: a column: 24

High quality sequence start: 29

High quality sequence stop: 666.

FEATURES

Location/Qualifiers

1..1076

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5574047"

/issue\_type="duodenal adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH MGC 88"

/note="Organ: small intestine; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."

BASE COUNT 215 a 357 c 310 g 193 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 5,53e-200 Length: 1076

Score: 235.00 Matches: 275

Percent Similarity: 99.28% Conservative: 0

Best Local Similarity: 99.28% Mismatches: 1

Query Match: 8.04% Indels: 2

DB: 12 Gaps: 0

US-09-916-849a-3 (1-2923) x BM480097 (1-1076)

QY 1942 GlnCYsProCYsLYrProGlyValIleGlyArgGlnCYsAspArgCYsAspAsnProPhe 1961

DB 38 CAGGTGTCAAGCAAGCAAGGTGTATCGGCGTCAAGTGTACCGCTGTGACACCTTTT 97

QY 1962 AlaGluValThrThrAsnGlyCYsGlnValAsnTYrAspSerCYsProArgAlaIleGlu 1981

DB 98 GCTAAGGTCAACCAATGCTGTGAAGTATATGACAGCTGCCACGACGACATTGAG 157

QY 1982 AlaGlyIleTPRProArgTYrThrArgPheGlyLeuProAlaAlaIleProCYsProLYs 2001

DB 158 GCTGGAACTGTGTGGCCCCGTAACCGCTTGGGCTGTGCTGTCTCTCCCTGCCAAA 217

QY 2002 GlySerPheGlyThrAlaValArgHisCYsAspGlnHisArgGlyTYrLeuProProAsn 2021

DB 218 GGCCTCTTGGAGTGTCTGTGCCCACTGTGATAGACACAGGGGTGGCTCCCCCAAC 277

QY 2022 LeuPheAsnCYsThrSerIleThrPheSerGlnLeuLYrGlyPheAlaGluArgLeuGln 2041

DB 278 CTCTTCAACTGCACAGTCCATCACTTCTCAGAACTGAAGGGCTTCGCTGAGCGGCTACAG 337

QY 2042 ArgAsnGlySerGlyLeuAspSerGlyLYrArgSerGlnGlnLeuAlaLeuLeuLeuArgAsn 2061

DB 338 CGGAATAGTCAAGGCTCTAGACTCAGGGCGCTCCAGACGCTAGGCTGTCTTCGGCAAC 397

QY 2062 AlaThrGlnHisThrAlaGlyTYrPheGlySerAspValLYrValAlaTYrGlnLeuAla 2081

DB 398 GCGACGAGGACACAGCTGGCTACTTGGGACGACGACCTCAAGGTGGCTACAGCTGGCC 457

QY 2082 ThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAsp 2101

DB 458 ACGCGGCTGTGGGCCACAGAGACACCCAGCGGGCTTGGGCTGTGCTGCCACACAGAC 517

QY 2102 ValHisPheThrGlnValAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsnLYs 2121

DB 518 GTGCATTTCAAGAAATCTGTGCGGGTGGGAGCGCCCTCTGTGACACAGCCACAG 577

QY	2122	ArgHisrPbGluLeuLeuIleGIInthrGluGluYgIYrAlaIaTrpLeuLeuGIHisrYr	2141
Db	578	CGGCACTGGAGCTGATCTCCAGCAGACAGAGGGTGGCACCCGCTGCTCTCCAGACATAT	637
QY	2142	GIuAlaYrAlaSerAlaLeuAlaGlnaMetArGHisrYrLeuSerProPheThr	2161
Db	638	GAGGCGTACGGCCAGGCCCTGGCCCGAGAACATGCGCGACACCTACCTAGGCCCTTCACCC	697
QY	2162	IleValThrProaenIleValIleSerValValArgLeuAsp-LysGluYaenPheAlaGI	2181
Db	698	ATTCGTCAAGCCCAACTTGTCTCATCTCCGTAGGGCGCTTGGACAAAGG-AACTTTGCTGG	756
QY	2181	YAlaIaYsLeuProArGTYrGluAlaLeuArGluYgIuGlnProPoaPLeuGIuThrTh	2201
Db	757	GGCCAGAGCTGCCCGCTACGAGAGCCCTCGTGGGAGAGAGGCCCGCGAGCCTTGAGACAAC	816
QY	2201	rValIleLeuProGluSerValaPheArGluThrProProValValArg	2217
Db	817	AGTCAATTGCTCTAGATCTGTCTTTTCAGAGAGAGAGCCCGCCGCTGTGTCAAG	865

LOCUS	BUS57213	908 bp	mRNA	linear	EST 16-OCT-2002
DEFINITION	AGENCOURT 10480787 NTH_MGC_107 Homo sapiens cDNA clone IMAGE:6647022 5', mRNA_sequence.				

ACCESSION	B0657213	GI:24042203
VERSION	B0657213.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 908)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished
CONTACT	Contact: Robert Strausberg, Ph.D.

Email: [cs9apbs-r@mail.nih.gov](mailto:cs9apbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
plate: LICM2886 row: m column: 06  
High quality sequence stop:646.

FEATURES	Location/Qualifiers
source	1. .908

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="6647022"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH MGC 107"
/note="Organ: breast; Vector: pOT8; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAC(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC Library."

```

	BASE COUNT	190 A	300 C	281 G	137 T
ORIGIN					
Alignment Scores:					
Pred. No.:		5.83e-192		length:	908
Score:		226.00		Matches:	226
Percent Similarity:		100.00%		Conservative:	0
Best Local Similarity:		100.00%		Mismatches:	0

Query Match:	7.73%	Indels:	0
DB:	13	Gaps:	0
US-09-916-849A-3 (1-2923) x BU857213 (1-908)			

Qy 2654 ArgLeuTyrGlnProTyrGlnYaspSerAlaGlySerLeuHISerThrSerArgSerGly 2673  
Db 2 CGGCTGACAGGCTTACGGAGACTCGGCTCTCTGCACAGCACCAAGTCGGCTCGGAC 61

QY 2674 LysSerGlnProSerTyrIleProPheLeuAArgIuGluSerAlaLeuAsnProIly 2693

Dp 62 AAGAGTCAAGCCACGCTACATCCCTTCTTGCTGAGGGAGAGTCCGCACTGAACCTG9C 121

Oy 2694 GINGLYPPTFCGLYLENGLYASPPTFCGLYSERLEUHEULENGLYGLINASPGLND 2113  
 Db 122 CAAGGGCCCCCTGGCTGGCGGATCCAGGCAGCTGTTCCTGGAAAGCTCAAGACACAGCAG 181

Oy 27/4 K18ASPpROASPnhtasPseraspSeaspLeuseitEugluaspAspGintErglyser 2/53  
 Db 182 CATGATCTGACACGGACCTCCGACAGTACCTCTTGAAGAAGACGACCAAGAGTGCTCC 241

**Db**

2754 TATGACCTTACCACACTCATCAGACAGTGAGAGAGAGAGAGAGGCC 301

Db

302 GCGTTCCTGGAGAGCAGGCGCTGGGAATAGCCGTCTGGGCGCTGGAGACGAGAGA CTGCC 361

362 CTGCACACTACTCCAGAGATGGGGGCCCGAGGCCTGCGAAGGCCCTTGSCCAGGAGAC 421

422 TTTCGGACCAACACAAAAGAGACTAGTCGCAACGGGCGCCCTCAGAGACGGCGTCGGCGAG 481

482 AATGAGATGCCCTGTCTCGAAGAGGGGTCCTTAGGCCCCCTTCCAGGCTCTTCTGCCAG 541

Db 542 CCGCACAAAGGCATCTTAAGAAAGAGTGTCTGCCCCACCATCAGCGAAGACAGCAGCCTC 601

Db	602	CTCCGCGCTCCCTCGAGCAATGCACAGGGTCTTCCCGGGGCTCTCCGCTAGTGAAGGC	661
Ov	2874	SeFArgrlvglvProPro	2879

Db 662 AGCGGGGCGGGCCCCC 679

Accession	Gene	Size	Library	Strain	Species	Library	Strain	Species
AF116945	LOCUS							
AF116945	DEFINITION	695 bp	mRNA	linear	EST	01-AUG-2002		
AF116945	HEMBA1							
AF116945	Homo sapiens	cdna	clone	HEMBA1000330	5'	mRNA		
	sequence.							

```

VERSION      AUT16945.1  GI:10931871
KEYWORDS
SOURCE
EST.         Homo sapiens (human)

```

REFERENCES  
Bukacynska, M.: 1965, 'Chordata, Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. I (pages 1 to 695)'.

TITLE Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
Isogai, T.  
HRI human CDNA project  
1000000

COMMENT  
Contact: Takao Ito  
Genomics Laboratory

Helix Research Institute  
1532-3 Yana, Kibaraazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

# FEATURES

## source

1.695  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEMBA1000330"  
/issue\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/clone\_1ib="HEMBA1"  
/note="Vector: pME18SFL3"  
BASE COUNT 105 a 246 c 174 g 167 t 3 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.1e-177 Length: 695  
Score: 210.00 Matches: 210  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.18% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AUI16945 (1-695)

QY 2399 ThrLeuLeuArgIleLeuArgSerAenGlnHsGlyIleArgArgAsnLeuThraAla 2418  
DB 1 ACTCTCTTCGATCTCGTGCCTCAACCAACAGGATCGAGTAACTGACAGCTGCG 60  
QY 2419 LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAenGlnAlaAspLeuProPheAla 2438  
DB 61 CTGGGCGCTGGCTCAGTGGTCTCTCTCGGGAATCAACAGGCTGACCTCTTCTTGGC 120  
QY 2439 CysThrValIleAlaIleLeuLeuHisPheLeuTyrlLeuCysThrPheSerTrpAlaLeu 2458  
DB 121 TGCAACAGTATGCGCATCTGCTGCACTTCCTGACCTTCTTCTGGGCTCTG 180  
QY 2459 LeuGlnAlaLeuHisLeuTyrlArgAlaLeuThrgIleValArgAspValAsnThrgIlePro 2478  
DB 181 CTGGAGGCGCTTCACCTGACCGGGCACTCAGTGAAGTGCAGATGCAACCGGCC 240  
QY 2479 MetArgPheTyrlMetLeuGlyTrpGlyValProAlaPheIleThrgIleuAlaVal 2498  
DB 241 ATGCCCTCTTACTACATGCTGGGCTGGGCGTGGCTGCTTCATCAACAGGCTGACCGCTG 300  
QY 2499 GlyLeuAspProGlnGlyTyrlGlyAsnProAspPheCysTrpLeuSerIleTyrlAspThr 2518  
DB 301 GGCCCTGAGACCCGAGGGCTACCGGAACTTGACTTGTGCTCTCCATCTATGACAG 360  
QY 2519 LeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrl 2538  
DB 361 CTCACTCTGAGATTGTTGCTGGCGGCTGGCTTGGCGCTTCGATGAGTCTTCTTCGTAC 420  
QY 2539 IleLeuAlaAlaArgAlaSerCysAlaAlaGlnGlnGlyPheGlnIleTyrlGlyPro 2558  
DB 421 ATCTGGCGGCGCGGCGCTCTCTGTGTGCTGCCAGCGGCAAGGCTTGAAGAAAGTCTCT 480  
QY 2559 ValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeu 2578  
DB 481 GTCCTGGGCGCTGACGCGCTCTTGGCGCTCTCTGTGCTGAGGCCACGCGGCTGCTG 540  
QY 2579 AlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrlLeuPheAlaThrCysAsn 2598  
DB 541 GCACGTCTCTCTGTCAACAGCAACCCCTCTTCCACTACTCTTGTGCTACTGCAAT 600  
QY 2599 CysIleGlnGlyProPheIlePheLeuSer 2608

DB 601 TGCACTCAGGCGCTTCATCTCTCTCTCC 630

## RESULT 6

AL042580/c 662 bp mRNA linear EST 29-FEB-2000  
LOCUS DKF2p34K0121.r1 434 (synonym: htes3) Homo sapiens cDNA clone  
DEFINITION DKF2p34K0121 5', mRNA sequence.

ACCESSION AL042580  
VERSION AL042580.1 GI:5422032  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 662)  
AUTHORS Blum H., Baurer-Sch., S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
TITLE EST (Blum, et al.)  
JOURNAL Unpublished  
COMMENT Unpublished  
CONTACT: Blum H  
MIPS

Am Klopferplatz 18a D-8152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by LMU (Ludwig Maximilians University,  
Munich/Germany) within the cDNA sequencing consortium of the German  
Genome Project.

No al sequence available.  
This clone (DKF2p34K0121) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

### source

1.662  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKF2p34K0121"  
/issue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="MD10B"  
/clone\_1ib="434 (synonym: htes3)"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"  
BASE COUNT 148 a 197 c 179 g 138 t  
ORIGIN

## Alignment Scores:

Pred. No.: 5.24e-175 Length: 662  
Score: 207.00 Matches: 207  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.08% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AL042580 (1-662)

QY 380 SerThrThraAlaValPheLeuSerValGluAspAsnAspAlaProGlnPhe 399  
DB 623 AGTACCAAGCGCGTGTCTTCTTCTGTGAGAGATGACATGATATGCCCCCAAGTTT 564  
QY 400 SerGlnGlyAspGlyValAlaGlnValArgGlnAspValThrProGlyAlaProValLeu 419  
DB 563 AGTGAAGAGCCCTATGAGTCCAGTGAAGGAGATGTGACTCCAGGGGCCCCAGTACTC 504  
QY 420 ArgValThraAspAspAspAspGlySerAsnAlaValValHisTyrlSerIleMet 439  
DB 503 CGAGTCAAGCTCTCGATCGAGACAGAGGAGCAATGCCGAGTGCATATGATCATG 444  
QY 440 SerGlnAsnAlaArgGlyGlnPheTyrlLeuAspAlaGlnThrgIleValAlaAspValVal 459  
DB 443 AGTGGCAATGCTCGGGGACAGTTTATCTGAGATGCCAGACTGAGCTCTGATGTGGTG 384  
QY 460 SerProLeuAspTyrlGlnThrThrgIleGlnIleTyrlThrgIleuArgValAlaGlnAspGly 479



```

Db      383 AGCCCTCTTACTATGAGACACCAAGAGTACACCTTACGGGGTCAGACACAGATGGT 324
Qy      480 G1YXgProProleuSerAanValSerG1YleuValThrValGlnValLeuAapIleAan 439
Db      323 GGCCTGCTCCCATCTTAAATGCTCTGGCTTGGTGAAGTACAGGCTCGAATATCAAC 264
Qy      500 AAspAaIaProIlePheValSerThrProPheGlnA1aThrValLeuG1uSerValPro 519
Db      263 GACAAATGCCCACTTCTGTCAGACACCTTTCCAGCTACTGTCCTCGAGAGAGTCCCC 204
Qy      520 LeuG1YTrileuValLeuH1SvalGlnAlaIleAapIleAapIleAapIleAapIleAap 539
Db      203 TTAGGCTACCTGCTTCTCCATGTCAGGCTATGACCTGATCTGCTGCAATATGCCGC 144
Qy      540 LeuG1YTrileuValLeuH1SvalGlnAlaIleAapIleAapIleAapIleAapIleAap 559
Db      143 CTGGAAATACCGCTTCTGCTGAGGAGGACATGACTTCCCTTCCATCAACATAGGACA 84
Qy      560 G1YTrileuValLeuH1SvalGlnAlaIleAapIleAapIleAapIleAapIleAap 579
Db      83 GGCTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24
Qy      580 ValG1uA1aArgAapIleGly 586
Db      23 GTAGAAGCTCGAGACCATGGC 3

RESULT 7
BX409741 877 bp mRNA linear EST 13-MAY-2003
LOCUS    BX409741 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0D013YE18 5-PRIME, mRNA sequence.
ACCESSION BX409741
VERSION   BX409741.1 GI:30652999
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Bunkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 877)
          Li, W.B., Gruber, C., Jesses, J. and Polyes, D.
          Full-length cDNA libraries and normalization
          Unpublished
          Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 64.f For more
          information about this cluster, see http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CS0BAF012ZG07.AF01112.1&cluster=64.f.
          Contact : Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0BAF012ZG07.AF01112_1.
FEATURES
     source
         1..877
             location/Qualifiers
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="CS0D013YE18"
                 /tissue_type="FETAL BRAIN"
                 /dev_stage="Fetal"
                 /clone_lib="Homo sapiens FETAL BRAIN"
                 /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
                 was primed with a NotI-oligo(dT) primer. Five prime end
                 enriched, double-strand cDNA was digested with Not I and
                 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                 vector. Library was not normalized."
BASE COUNT      140 a      315 c      226 g      194 t      2 others
ORIGIN
Alignment Scores:      3.31e-172      Length:      877
Pred. No.:

```

```

Score:      204.00      Matches:      204
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      6.98%      Indels:      0
DB:      13      Gaps:      0

US-09-916-849A-3 (1-2923) x BX409741 (1-877)
Qy      2441 ValIleAlaIleLeuLeuH1SvalGlnAlaIleAapIleAapIleAapIleAapIleAap 2460
Db      29 GTCATTTGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
Qy      2461 AlaLeuH1SvalGlnAlaIleAapIleAapIleAapIleAapIleAapIleAap 2480
Db      89 GCTTGACCTGTAACCGGCACTCACTAGAGGCGGAGTCAACACCGGCCCATGGCC 148
Qy      2481 PheYTrMetLeuG1YTrG1YValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500
Db      149 TTCTACTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
Qy      2501 AAspProG1uG1YTrG1YValProAapPheCy8TrPleuSerIleTyraPheThrLeu 2520
Db      209 GACCCGAGGAGGCTACGGGAACTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 268
Qy      2521 TrPserPheAlaG1YProValAlaPheAlaValSerMetSerValPheLeuYTrileu 2540
Db      269 TCGAGTTTGTGCTGGCCCGGCTGCTTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 328
Qy      2541 AlaAlaAgaIleSerCy8AlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2560
Db      329 GGGGCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388
Qy      2561 G1YLeuGlnProSerPheAlaValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 2580
Db      389 GGCTGACAGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 448
Qy      2581 LeuSerValAanSerAapThrLeuLeuPheH1SvalGlnAlaIleAapIleAapIleAap 2600
Db      449 CTCTGTCACAGAGACACCTCTCTCTCCATCTTGTGCTACCTGCAATGGCATTC 508
Qy      2601 GlnGlyProPheIlePheLeuSerTyTrValValLeuSerTyGlnValArgValAlaLeu 2620
Db      509 CAGGCGCCCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 568
Qy      2621 LysLeuAlaCySerAryLysProSerProAapProAlaLeuThrThrLysSerThrLeu 2640
Db      569 AAGCTTGCTGAGCGCGAGCGAGCCCTTACCTGCTGTCAGACCAAGTCCACCTCG 628
Qy      2641 ThrSerSerTyTr 2644
Db      629 ACCTGCTCTAC 640

RESULT 8
BQ677400 906 bp mRNA linear EST 15-JUL-2002
LOCUS    BQ677400 AGENCOURT 8209038 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260299
DEFINITION 5', mRNA sequence.
ACCESSION BQ677400
VERSION   BQ677400.1 GI:21790079
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Bunkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 906)
          NIH-MGC http://mgi.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: DCTD/DRP
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (UMNL)

```

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>  
 plate: LLCM2419 row: c column: 20  
 High quality sequence stop: 606.

## FEATURES

source

1..906  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6260299"  
 /tissue\_type="melanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 112"  
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."  
 BASF COUNT 132 a 366 c 226 g 169 t 13 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.06e-167 Length: 906  
 Score: 199.00 Matches: 199  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.81% Indels: 0  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BQ677400 (1-906)

QY 2675 SerGlnProSerTyrTlleProPheLeuAArgGluGluSerAlaLeuAnProGlyGln 2694  
 DB 3 AGTCAGCCCAAGCTACATCCCTCTTCTGAGGAGAGACTCCGCACTGAACCTTGCCCA 62  
 QY 2695 GlyProGlyLeuGlyAAspProGlySerLeuPheLeuGluGlyGlnAAspGlnHis 2714  
 DB 63 GGGCCCCCTGGCTGGGGGATCCAGGCAAGCTGTTCTTGAGAGCTCAAGACAGACAT 122  
 QY 2715 AspProAAspThrAAspSerAAspSerAAspLeuSerLeuGluAAspAAspGlnSerGlyTyr 2734  
 DB 123 GATCCTGACACGAGACTCCGACAGTGAACCTCTTAAGAGACAGACAGAGTGGCTCAT 182  
 QY 2735 AlaSerThrHisSerSerAAspSerGluGluGluGluGluGluGluGluGluGluGlu 2754  
 DB 183 GCTCTACCCCACTCATCAAGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242  
 QY 2755 PheProGlyGluGluGlyTTPAAspSerLeuLeuGlyProGlyAlaGluAArgLeuProLeu 2774  
 DB 243 TTCCCTGAG 302  
 QY 2775 HisSerThrProGlyAAspGlyGlyProGlyProGlyProGlyProGlyProGlyPro 2794  
 DB 303 CACACTACTCCCAAGAGATGGGGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362  
 QY 2795 GlyThrThrAlaLeuGluSerSerGlyAAspGlyAlaProGluGluAArgLeuAAspGlu 2814  
 DB 363 GGAAGACACAGCAAAAGAT 422  
 QY 2815 GlyAAspAlaLeuSerAAspGlySerLeuGlyProLeuProGlySerSerAlaGlnPro 2834  
 DB 423 GGAAGATGCCCTGTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 482  
 QY 2835 HisLeuGlyTlleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 2854  
 DB 483 CACAAAGGATCTCTTAAGCT 542  
 QY 2855 ArgLeuProLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2873

DB 543 CGGCTCCCTTGAGAGATGACAGAGGCTTCTCCGGGGCTCTCGCTAAGTGAAGGC 599

## RESULT 9

BQ679970

LOCUS

DEFINITION

BOG79970

AGENCOURT 8212562 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6260781

5', mRNA sequence.

ACCESSION

BOG79970

VERSION

BOG79970.1 GI:21792649

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 927)

NIH-MGC <http://imgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNL at:

<http://image.llnl.gov>

plate: LLCM2420 row: g column: 22

High quality sequence stop: 612.

Location/Qualifiers

1..927

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6260781"

/tissue\_type="melanotic melanoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 112"

/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."

BASF COUNT 133 a 373 c 319 g 92 t 10 others

ORIGIN

## Alignment Scores:

Pred. No.: 1.66e-160 Length: 927  
 Score: 191.00 Matches: 204  
 Percent Similarity: 99.51% Conservative: 0  
 Best Local Similarity: 99.51% Mismatches: 0  
 Query Match: 6.53% Indels: 1  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BQ679970 (1-927)

QY 2675 SerGlnProSerTyrTlleProPheLeuAArgGluGluSerAlaLeuAnProGlyGln 2694  
 DB 3 AGTCAGCCCAAGCTACATCCCTCTTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 62  
 QY 2695 GlyProGlyLeuGlyAAspProGlySerLeuPheLeuGluGlyGlnAAspGlnHis 2714  
 DB 63 GGGCCCCCTGGCTGGGGGATCCAGGCAAGCTGTTCTTGAGAGCTCAAGACAGACAT 122  
 QY 2715 AspProAAspThrAAspSerAAspSerAAspLeuSerLeuGluAAspAAspGlnSerGlyTyr 2734  
 DB 123 GATCCTGACACGAGACTCCGACAGTGAACCTCTTAAGAGACAGACAGAGTGGCTCAT 182  
 QY 2735 AlaSerThrHisSerSerAAspSerGluGluGluGluGluGluGluGluGluGluGlu 2754

Db		183	GCGCTACCCACTCATCAGACAGTGGAGGAAAGAAAGCAGAGGAAGAAGAGGCCGCC	242
Oy		2755	PheProglYgIugInGlYTTPApSeriLeuEnGlYProGlYAlaGlUArgLeuProLeu	2774
Db		243	TTCCCTGGAGAGCAGGCCTGGAGTAAGCTGCTGGGGCCTGGAGCAGAGACTGCCCTG	302
Oy		2775	HISerThrPOLYAsPGLYgIYProGlYPProGlYLYsAlAPoTrPProGlYAsPHe	2794
Db		303	CACAGTACTCCCAAGAGATGGGGGCCACAGGCTGGCAAGCCCCCTGGCCAGAGACTTT	362
Oy		2795	GIYThrThraLYsGLIsErSeRcLYsAngLYAlaProGlUGluArgGLuaSn	2814
Db		363	GGGACCACAGAAAAAGAGTAGTGTCACCGGGCCCTGAGAGCGGCTCCGGAGAT	422
Oy		2815	GLYsPaLaLeu-SerArGJUGlYSerLeuGlYProLeuProGlYSerSerAlaINr	2834
Db		423	GGAGTGGCTTNGTCTCGAAGGGGTCCCTTAgGCCCTTCAGGCTTTTGCCAGCC	482
Oy		2834	OHIsLySGlyLleLeuLYsLYsCYsLeuProThrLleserGUlysSerSerLeuE	2854
Db		483	TCACAAAGCATCTTAGAAGAAAGTGTCTGCCCAATCAGCGAAGAGACAGACTCTCT	542
Oy		2854	WArgLeuProLeuGInGlucYethrGLysSerArGlySerSerAlaserglYsE	2874
Db		543	GCGGCTCCCCCTGGAGCAATGACAAGGCTCTCCGGGGCTCTCCCTAGTAGAGGCGAG	602
Oy		2874	rArgLYgLYPro 2878	
Db		603	CCGGGGCGGNCCTG 615	
RESULT 10				
LOCUS	B0676596			
DEFINITION	AGENCOURT 8196790 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6263505	930 bp	mRNA	linear EST 15-JUL-2002
ACCESSION	B0676596			
VERSION	B0676596.1	GI:21789275		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiide; Homo.			
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/.			
AUTHORS	1 (bases 1 to 930)			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCD/DNP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LRCM2427 row: 1 column: 10 High quality sequence scop: 602. Location/Qualifiers 1..930 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6263505" /tissue_type="melanotic melanoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 112" /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGACAGAG(G). library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthase kit (Stratagene) and			
FEATURES				
source				

[illegible]

TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-riemail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM2828 row: h column: 17  
High quality sequence stop: 556.  
Location/Qualifiers  
1. 893  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGB:6597377"  
/issue\_type="teratocarcinoma, cell line"  
/lab\_host="PH10B (phage-resistant)"  
/clone\_id="NIH MGC 109"  
/note="Organ: ovary; Vector: pOTB7; Site 1: SCOR1; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into SCOR1/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

BASE COUNT 179 a 268 c 274 g 172 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6,32e-157 Length: 893  
Score: 187.00 Matches: 187  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.40% Indels: 0  
DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BUB49091 (1-893)

QY 1689 LeuGluProGlyAlaArgAlaAsnArgGlyAAspTPhIshAlaGlnLeuAlaLeuGlyAla 1708  
DB 2 CTGGAGCCAGCGCCGCGCCATGACGGTGACTGGACACATGACAGCTGGACCTGGAGCC 61

QY 1709 SerGlyGlyProGlyAlaAlaIleLeuSerPheAspTyrGlyGlnAlaArgAlaGly 1728  
DB 62 AGCGGGGGGGCCGGCCATCCATTCCTGCTTCGATTATGGGAGAGAGAGAGAGAGAG 121

QY 1729 AsnLeuGlyProArgLeuHISgIleuHISLeuSerAsnIleThrValGlyGlyLeuPro 1748  
DB 122 AACCTGGGGCCCCGGCTGCATGCTGCACCTGAGCAATAACTGGGGGGAATACCT 181

QY 1749 GlyProAlaGlyGlyValAlaArgGlyPheArgGlyGlyLeuGlnGlyValAlaSer 1768  
DB 182 GGGCCAGCGCGGGGCTGAGCCGCTTCCTGGGGCTGTTTCAGGGTGTGGGGTGGAGC 241

QY 1769 AspThrProGlyGlyValAlaSerLeuAspProSerHisGlyGlnSerIleAsnValGlu 1788  
DB 242 GATTACCCCGAGAGGGGCTTAACAGCTGATCCAGACCATGGGAGAGAGATCAACGTGAG 301

QY 1789 GlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCys 1808  
DB 302 CAGAGGTGAGCTGCTGACCTTGTGACTCAAAACCCGTGCTGCTAAGCATATTC 361

QY 1809 SerAsnAspTyrAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAspAsnCys 1828  
DB 362 AGCAACGACTGGAGACGACTATTCCTGACCTGTGATCCAGGTACTATAGTGAACAACGT 421

QY 1829 ThrAsnValCysAspLeuAsnProCysGlnISgInSerValCysThrArgIysProSer 1848

DB 422 ACTAATGTGTGTGACCTTAACCCGTGTGAGACACCACTGTGTGTGACCCGGAAGCCAGT 481

QY 1849 AlaProHISgIlyTyrThrCysGluCysProProAsnTyrLeuGlyProTyrCysGluThr 1868  
DB 482 GCGCCCATGGCTAATACCTGGAGAGTGTCCCAATTAATCTTGGGCCATATCTGTAGAC 541

QY 1869 ArgIleAspGlnProCysPro 1875  
DB 542 AGGATTTGACCAAGCTTGTGCC 562

RESULT 12  
AW370778  
LOCUS AW370778 640 bp mRNA linear EST 04-FEB-2000  
DEFINITION QV0-BT0263-101299-072-G04 BT0263 Homo sapiens CDNA, mRNA sequence.  
ACCESSION AW370778  
VERSION AW370778.1 GI:6875432  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE HCCP http://www.ludwig.org.br/ORBSTS.  
AUTHORS The FAPESP/LICR Human Cancer Genome Project  
TITLE Unpublished  
JOURNAL Contact: Stimpson A.J.G.  
COMMENT Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL.  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV0&ct2-QV0-BT0263-101299-072-g04&ct3=1999-12-10&ct4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 640.  
Location/Qualifiers  
1. 640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_id="Br0263"  
/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORBSTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 122 a 213 c 160 g 145 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7,32e-150 Length: 640  
Score: 179.00 Matches: 207  
Percent Similarity: 99.04% Conservative: 0  
Best Local Similarity: 99.04% Mismatches: 1  
Query Match: 6.12% Indels: 2  
DB: 9 Gaps: 0

US-09-916-849A-3 (1-2923) x AW370778 (1-640)

QY 2247 AlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHISAsnTyrAspProAsp 2266  
DB 15 GCGAGCGTATCATCTAACCGACCTGGCCGCGGCTACTGCTCATTAATGACCTTGC 74

QY 2267 LysArgSerLeuAspValProLysArg-ProIleIleAsnThrProValValSerIleSe 2286  
DB 75 AAGGCGAGCTTGAAGAGTCCCAATG-CCGATATATCAACACCCCGTGGAGCATCAG 133



SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 523)  
 Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,  
 Radetof, U., Schneider, D. and Korn, B.  
 TITLE Human Unigeneset - RZPD3  
 JOURNAL Unpublished  
 COMMENT Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
 RZPD; IMA9998K04177.  
 RZPDLIB; I.M.A.G.E. CDNA Clone Collection;  
 Human Unigeneset - RZPD3 (RZPDLIB No. 972)  
 http://www.rzpd.de/CloneCards/cgi-  
 bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de

FEATURES  
 SOURCE This clone is available royalty-free from RZPD;  
 contact RZPD (clone@rzpd.de) for further information. Seq primer:  
 M13u, Primer sequence: CGTGTAAACGACGCCAGT.  
 Location/Qualifiers  
 1..523

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAG998K04177 ; IMAGE:42919"  
 /sex="female"  
 /dev\_stage="73 days post natal"  
 /lab\_host="PH10B (ampicillin resistant)"  
 /clone\_lib="Soares infant brain INTB"  
 /note="Organ: whole brain; Vector: lacmid BA; Site 1: Not  
 1, Site 2: Hind III; 1st strand cDNA was primed with a Not  
 1 - oligo(dT) primer [5',  
 AACTGAGAGATTCGGCGCGGACGAGATTTTCTTTTCTTTT 3'];  
 double-stranded cDNA was ligated to Hind III adaptors  
 (Pharmacia), digested with Not I and directionally cloned  
 into the Not I and Hind III sites of the lacmid BA vector.  
 Library went through one round of normalization. Library  
 constructed by Bento Soares and M. Patricia Bonaldo."  
 BASE COUNT 109 a 183 c 133 g 97 t 1 others

Alignment Scores:  
 Pred. No.: 2,95e-138 Length: 523  
 Score: 166.00 Matches: 166  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.68% Indels: 0  
 DB: 13 Gaps: 0

US-09-916-849A-3 (1-2923) x BK110290 (1-523)

QY 2585 SerAapThLeuLeuPheHisTyrLeuPheAlaThrCysAaNCysAlIeGInIyProPhe 2604  
 DB 26 AGCGACACCCCTCTTCCACTACCTCTTGGTACTGCAATTCGATCCAGGCCCCCTTC 85  
 QY 2605 IlePheLeuSerTyrValIleuSerIyrgIValIyAlaIyAlaIyAlaIyAlaIy 2624  
 DB 86 ATCTTCCTCTCTAATGTGTGTACAGAGAGGTCCGAAAGCACTCAAGCTTCGCTGC 145  
 QY 2625 SerATgIyPProSerProAapProAlaLeuThrThIySerThIeuthIserSerTyr 2644  
 DB 146 ACCCGAAGCCAGCCCTGCTGTGACACCAATCCACCTGACCTGCTGCTTAC 205  
 QY 2645 AaNCyPProSerProTyrAlaAapGlyAglLeuTyrGInProTyrGlyAaPserAlaGly 2664  
 DB 206 AACTGCCCAAGCCCTTACCGACATGGCGGCTGTACCAAGCTTACCGAGACTCTGCGCGGC 265

QY 2665 SerLeuHIsSerThSerAysSerGlyIySerserInProSerTyrIleProPheLeuLeu 2684  
 DB 266 TCTTCGACACGACACCGATCGCTCGGGCAAGTCAGCCAGCTACATCCCTTCTTGGCTG 325  
 QY 2685 ArgGIuGIuSerAlaLeuAaProGIuGIuInGIyProProGIuGIuAaPProGIuYer 2704  
 DB 326 AGGAGAGAGTCCGACCTGAACCTTGGCCAGAGGCCCTTGGGCTGGGATCCAGCGAGC 385  
 QY 2705 LeuPheLeuGIuGIuGIuAaPInGIuHIsaPProAapThIaPserAaPserAaPleu 2724  
 DB 386 CTGTTCTCGAAGAGTCAGACAGCAGCATATCTTGACAGGACTCCGACAGTACCTCG 445  
 QY 2725 SerLeuGIuAaPAPGIuSerGIySerTyrAlaSerThIeSerSerAaPserGIuGIu 2744  
 DB 446 TCTTGAAGAGCGACCAAGTGTCTCTATGCTCTACCACTCAGACAGTGAAGG 505  
 QY 2745 GIuGIuGIuGIuGIuGIuGIu 2750  
 DB 506 GAAAGAGAGAGAGAGAA 523

RESULT 15  
 BF951925/c 549 bp mRNA linear EST 22-JAN-2001  
 LOCUS OVI-NN0228-091100-440-f05 NN0228 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BF951925  
 ACCESSION BF951925.1 GI:12369200  
 VERSION BF951925.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 549)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P.,  
 Goldman, G.H., Carvalho, A.F., Matsumura, A., Baita, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.V. and  
 Simpson, A.J.U.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 JOURNAL 20202663  
 MEDLINE 10737800  
 PUBMED Contact: Simpson A.J.U.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/getchml2.pl?tl=QV1-NN0228-  
 091100-440-f05&f3=2000-11-09&f4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 4  
 High quality sequence stop: 549.

#### FEATURES

##### source

1..549  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"

/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORSITES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 113 a 173 c 158 g 105 t

## ORIGIN

## Alignment Scores:

Pred. No.:	2,44e-137	Length:	549
Score:	165.00	Matches:	165
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.64%	Indels:	0
DB:	10	Gaps:	0

US-09-916-849a-3 (1-2923) x BP951925 (1-549)

```

QY      1707  GAlaSerGlyProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAla 1726
           |||
DB      547  GAGGCCAGGGGGGGGGCTGGCCATGCCATTCTGCTTGATATGGGACGAGAGACA 488
           |||
QY      1727  GAluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyGly 1746
           |||
DB      487  GAGGGCAACCTGGGGCCCCCGGCTGCATGCTCTGCACCTGCAGCAATAAACAGTGGGGCGGA 428
           |||
QY      1747  IleProGlyProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArg 1766
           |||
DB      427  ATACCTGGGCGACCGCGGGGTGGCCGCTTGCTTGCGGGCTGTTGCAAGGTGTGCGG 368
           |||
QY      1767  ValSerAspThrProGlnGlyValAsnSerLeuAspProSerHisGlyGlnSerIleAsn 1786
           |||
DB      367  GTGAGCGATACGCCGGAGGGGGTTAACAGCTGGATCCAGCCATGGGGAGAGCATCAAC 308
           |||
QY      1787  ValGlnGlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSer 1806
           |||
DB      307  GTGGAGCAGAGGCTGTAGCTGCTGACCCCTGTGACTCAAAACCGTGTCTCTGCTAAACAGC 248
           |||
QY      1807  TyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAspProGlyTyrTyrGlyAsp 1826
           |||
DB      247  TATTGCACGACACACTGGGACACTATTCTTGCACTGTGATCAAGTTACTATGGTGAC 188
           |||
QY      1827  AsnCysThrAsnValCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLys 1846
           |||
DB      187  AACTGTACTATATGTGTGACTGAACCCGTGTGAGACCAAGTGTGTGTACCCGCAAG 128
           |||
QY      1847  ProSerAlaProHisGlyTyrThrCysGlnCysProProAsnTyrLeuGlnProTyrCys 1866
           |||
DB      127  CCAAGTGCCCCCATGGCTATACCTGGAGTGTCCCCCAATTACCTGGGGCATACTGT 68
           |||
QY      1867  GluThrArgIleAsp 1871
           |||
DB      67  GAGACCGAGATTGAC 53
           |||

```

Search completed: February 14, 2004, 21:11:24  
 Job time : 11573 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:57:15 ; Search time 66 Seconds

(Without alignments)  
9273.097 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 15545  
Sequence: 1 MRSPTGVLPPTPPPLLL.....AGTVDESSGSSEFLFNFLLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15545	100.0	2923	9 US-09-788-711A-4	Sequence 4, App1
2	15545	100.0	2923	11 US-09-916-849A-3	Sequence 3, App1
3	15545	100.0	2923	12 US-10-174-677-29	Sequence 2, App1
4	15545	100.0	2923	12 US-10-120-801-53	Sequence 53, App1
5	15545	100.0	2923	12 US-10-292-798-932	Sequence 932, App1
6	15545	100.0	2923	15 US-10-225-567A-524	Sequence 524, App1
7	15518.5	99.8	2956	9 US-09-788-711A-2	Sequence 2, App1
8	8974.5	57.7	3034	9 US-09-737-149-25	Sequence 25, App1
9	8974.5	57.7	3034	9 US-09-737-149-30	Sequence 30, App1
10	8974.5	57.7	3034	12 US-10-131-409-70	Sequence 70, App1
11	8974.5	57.7	3034	12 US-10-120-801-52	Sequence 52, App1
12	8974.5	57.7	3034	12 US-10-150-811-70	Sequence 70, App1
13	8754	56.3	3014	9 US-09-737-149-2	Sequence 2, App1
14	8754	56.3	3014	12 US-10-241-230-107	Sequence 107, App1
15	8754	56.3	3014	12 US-10-174-677-77	Sequence 77, App1

16	8754	56.3	3014	12 US-10-295-027-750	Sequence 750, App1
17	8754	56.3	3014	12 US-10-240-145-63	Sequence 63, App1
18	8754	56.3	3014	15 US-10-225-567A-444	Sequence 444, App1
19	8709	56.0	3028	12 US-10-131-409-14	Sequence 14, App1
20	8709	56.0	3028	12 US-10-131-409-69	Sequence 69, App1
21	8709	56.0	3028	12 US-10-150-811-14	Sequence 14, App1
22	8709	56.0	3028	12 US-10-150-811-14	Sequence 69, App1
23	7833.5	50.4	3312	15 US-10-150-811-69	Sequence 69, App1
24	7780.5	50.1	3313	15 US-10-225-567A-656	Sequence 29, App1
25	7763	49.9	3298	15 US-10-149-819-21	Sequence 21, App1
26	4951	31.8	1713	9 US-09-737-149-27	Sequence 27, App1
27	4817.5	31.0	3579	12 US-10-174-677-5	Sequence 5, App1
28	4817.5	31.0	3606	12 US-10-120-801-54	Sequence 54, App1
29	3641	23.4	717	10 US-09-925-300-1299	Sequence 1299, App1
30	2958	19.0	568	9 US-09-843-856-2	Sequence 2, App1
31	2937	18.9	568	12 US-10-264-237-2041	Sequence 2041, App1
32	2900.5	18.7	565	12 US-10-176-847-100	Sequence 100, App1
33	2491	16.0	2610	12 US-10-369-493-6553	Sequence 6553, App1
34	2491	16.0	2610	12 US-10-369-493-6554	Sequence 6554, App1
35	2448.5	15.8	646	12 US-10-017-161-1096	Sequence 1096, App1
36	1712.5	11.0	4961	12 US-10-114-153-64	Sequence 64, App1
37	1534.5	9.9	5147	12 US-10-174-677-4	Sequence 4, App1
38	1516.5	9.8	4590	12 US-09-970-844-21	Sequence 21, App1
39	1516.5	9.8	4590	12 US-10-295-027-1323	Sequence 1323, App1
40	1516.5	9.8	4590	15 US-10-160-758-13	Sequence 13, App1
41	1516.5	9.8	4590	15 US-10-160-758-14	Sequence 14, App1
42	1516.5	9.8	4590	15 US-10-060-036-157	Sequence 157, App1
43	1460	9.4	458	12 US-10-029-386-33121	Sequence 33121, App1
44	1419.5	9.1	3097	12 US-10-174-677-6	Sequence 6, App1
45	1402	9.0	4544	12 US-10-085-198-22	Sequence 22, App1

#### ALIGNMENTS

RESULT 1  
US-09-788-711A-4  
; Sequence 4, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamara Testa  
; TITLE OR INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788, 711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSQ for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-4

Query Match 100.0%; Score 15545; DB 9; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRSPTGVLPPTPPPLLLLLPPPLADGVPCRSLSRRSGSCAAMPWMLCP	60
DB	1	MRSPTGVLPPTPPPLLLLLPPPLADGVPCRSLSRRSGSCAAMPWMLCP	60
QY	61	SASINMTYSCRDRAGTELNGHLPVHHDGLRWCPSEBAHPLPPADGCPW	120
DB	61	SASINMTYSCRDRAGTELNGHLPVHHDGLRWCPSEBAHPLPPADGCPW	120
QY	121	GHSPOGKLTLPESHPCIKAPRLRCQSCIKLAQAPGLRAGERSPEESL	180
DB	121	GHSPOGKLTLPESHPCIKAPRLRCQSCIKLAQAPGLRAGERSPEESL	180
QY	181	FGPSYQATVPENDPATTPVASLPAIDPDGEAGRLTYMDALPDSNSNO	240
DB	181	FGPSYQATVPENDPATTPVASLPAIDPDGEAGRLTYMDALPDSNSNO	240

Db 181 POPSYOATVBPENQAPGTVPVASLRAIDPDGEGARLEYTMALFDSRSNOFPFLDPEVTGA 240  
QY VTAABEDRETKSTHYERVTAADHGMFRSALATLTITVTDNDHDVPFEOQYKESLAE 300  
Db 241 VTTAAEIDRETKSTHYERVTAADHGMFRSALATLTITVTDNDHDVPEEQYKESLAE 300  
QY NLEVGVEVLTVARATDGDAPPANAILVRLLEGSGSGSPBEVIDPRSGVIRTRGPVDBREV 360  
Db 301 NLEVGVEVLTVARATDGDAPPANAILVRLLEGSGSGSPBEVIDPRSGVIRTRGPVDBREV 360  
QY 361 ESQVLTVEASDQGRDDEPRSTTAAPLSTVEDDNDNAQPSBKRYVYQVEDVTPGAPVLR 420  
Db 361 ESQVLTVEASDQGRDDEPRSTTAAPLSTVEDDNDNAQPSBKRYVYQVEDVTPGAPVLR 420  
QY VTAASDDEKGSNAVVHYSIMSGNARGOFTYDAQTGALDVSPLDYETTKETTLTVARADGG 480  
Db 421 VTAASDDEKGSNAVVHYSIMSGNARGOFTYDAQTGALDVSPLDYETTKETTLTVARADGG 480  
QY 481 RPLPSNVSGLVTVQVLDINDNAPIFVSTPQATVLSBVPJGYLVLAHQAIDADAGNARL 540  
Db 481 RPLPSNVSGLVTVQVLDINDNAPIFVSTPQATVLSBVPJGYLVLAHQAIDADAGNARL 540  
QY 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVDFTSFGYBARDHGTPTATASVSVTV 600  
Db 541 EYRLAGVGHDPFTTINNGTGMISVAALDREVDFTSFGYBARDHGTPTATASVSVTV 600  
QY 601 LDVNDNNPFTQPEYTVRLMEDAAGTAVTSAVDADASHVITTYQTSQNTNRPSITS 660  
Db 601 LDVNDNNPFTQPEYTVRLMEDAAGTAVTSAVDADASHVITTYQTSQNTNRPSITS 660  
QY 661 OSQGGVLSIALPLDYKLERQYVLAVTASDGTRODTAQIVNVTDANTHRPVPQSSHYTV 720  
Db 661 OSQGGVLSIALPLDYKLERQYVLAVTASDGTRODTAQIVNVTDANTHRPVPQSSHYTV 720  
QY 721 VNEDRPACTVTVLSATDEDTGENARITYEMEDSIPOFRIIDAPQATVTOAEIDYEDQVS 780  
Db 721 VNEDRPACTVTVLSATDEDTGENARITYEMEDSIPOFRIIDAPQATVTOAEIDYEDQVS 780  
QY 781 YTLAITARNDNGIPQKSDPTTYLEILVNDVNDNAPOFLADSYQSGSYEDVPPFTSVLQISAT 840  
Db 781 YTLAITARNDNGIPQKSDPTTYLEILVNDVNDNAPOFLADSYQSGSYEDVPPFTSVLQISAT 840  
QY 841 DROSGLANGRYFTYFQGGDDGDPFVASTSGIYRTLRRLRENVAAQYVLAAYVNDKMP 900  
Db 841 DROSGLANGRYFTYFQGGDDGDPFVASTSGIYRTLRRLRENVAAQYVLAAYVNDKMP 900  
QY 901 ARTPEMETVTVLVNDNDPVPFEBODEPVPFEBENSPIGLAARVATADPDEGTNAQIMYOI 960  
Db 901 ARTPEMETVTVLVNDNDPVPFEBODEPVPFEBENSPIGLAARVATADPDEGTNAQIMYOI 960  
QY 961 VEGNIPFVQJDI FSGELTALVDLDYEDREPEYLVITQATSAPLVSRATVVRLLDRNDP 1020  
Db 961 VEGNIPFVQJDI FSGELTALVDLDYEDREPEYLVITQATSAPLVSRATVVRLLDRNDP 1020  
QY 1021 PVUGNEPFIILNNYVYTNRSSPFGGALIGRVAPADPDISLITYSPREGENEISVTLNASTG 1080  
Db 1021 PVUGNEPFIILNNYVYTNRSSPFGGALIGRVAPADPDISLITYSPREGENEISVTLNASTG 1080  
QY 1081 ELKLSRALDNNRPLEAIMSVLVSDGVSVTAQCALRTIITIDEMLTHTSITRLLEDSMPER 1140  
Db 1081 ELKLSRALDNNRPLEAIMSVLVSDGVSVTAQCALRTIITIDEMLTHTSITRLLEDSMPER 1140  
QY 1141 FLSPBLGLFTIOAAVAATLAPPDHVVENVQORDTAPAGHILNVLVSLVGDPGPGGPPFL 1200  
Db 1141 FLSPBLGLFTIOAAVAATLAPPDHVVENVQORDTAPAGHILNVLVSLVGDPGPGGPPFL 1200  
QY 1201 PSEBLOERLYINBELTALISAQRYLPRDDNICAREPENTMRCVSVLRPSSAPFLASS 1260  
Db 1201 PSEBLOERLYINBELTALISAQRYLPRDDNICAREPENTMRCVSVLRPSSAPFLASS 1260  
QY 1261 VLFPRIPHVGGALRCRCPGFTGDYCFTEVDLCYSRPGPHGRCSRGSGYTCICRDOYTG 1320  
Db 1261 VLFPRIPHVGGALRCRCPGFTGDYCFTEVDLCYSRPGPHGRCSRGSGYTCICRDOYTG 1320

QY 1321 BECEVSARSRCCTPGVCKNGGTCVNLVVGFKDCDPCSDGFBKPYCOVTRSPFASFTTF 1380  
Db 1321 BECEVSARSRCCTPGVCKNGGTCVNLVVGFKDCDPCSDGFBKPYCOVTRSPFASFTTF 1380  
QY 1381 RGLRORFHTTALSPATYERDGLLLYNGRFNEKHDVLALEVIQBOVOLTFSAGESTTVS 1440  
Db 1381 RGLRORFHTTALSPATYERDGLLLYNGRFNEKHDVLALEVIQBOVOLTFSAGESTTVS 1440  
QY 1441 PVPVGSVDGQWHTVQKRYNPELIGQGLPGGSEBQKAAVTVVGCIDGVALRFSVIG 1500  
Db 1441 PVPVGSVDGQWHTVQKRYNPELIGQGLPGGSEBQKAAVTVVGCIDGVALRFSVIG 1500  
QY 1501 NYSCAAQGTQGGSKSLDTGLPBLIGVBDLSESPVEMRQFVGCGRNLQVDSRHIDMAD 1560  
Db 1501 NYSCAAQGTQGGSKSLDTGLPBLIGVBDLSESPVEMRQFVGCGRNLQVDSRHIDMAD 1560  
QY 1561 FIANNGTVPQCPAKKQVCDNSNTCHNGGTCVNQMDAPSCCEPLGFGKSCAQOMANPOHFL 1620  
Db 1561 FIANNGTVPQCPAKKQVCDNSNTCHNGGTCVNQMDAPSCCEPLGFGKSCAQOMANPOHFL 1620  
QY 1621 GSSLVAMHGLSLPISQPWTLIMFRTROADGVLLQATRGRETTILOAREGHVMTSVBGT 1680  
Db 1621 GSSLVAMHGLSLPISQPWTLIMFRTROADGVLLQATRGRETTILOAREGHVMTSVBGT 1680  
QY 1681 GLQASSLRLEPERANDGDWHHAQLALGASGGBGHAILSFQYQOQBAEGNLGRLHGLHS 1740  
Db 1681 GLQASSLRLEPERANDGDWHHAQLALGASGGBGHAILSFQYQOQBAEGNLGRLHGLHS 1740  
QY 1741 NITVGGIPBPAGVARGFRGCTQGYRVSDTPGVNSLDBSHGESINVEQGCSPDPCDSN 1800  
Db 1741 NITVGGIPBPAGVARGFRGCTQGYRVSDTPGVNSLDBSHGESINVEQGCSPDPCDSN 1800  
QY 1801 PCPANSYCSNDMDYSQSCDPEGYGDNCTNVCDLNPECOSQCTKPSAPHGTCCECPN 1860  
Db 1801 PCPANSYCSNDMDYSQSCDPEGYGDNCTNVCDLNPECOSQCTKPSAPHGTCCECPN 1860  
QY 1861 YLGPYCETRIDQPCRGMMWGHPTGCPNCNDVSKGFPDCKNTSGECHKENHYRPPGSP 1920  
Db 1861 YLGPYCETRIDQPCRGMMWGHPTGCPNCNDVSKGFPDCKNTSGECHKENHYRPPGSP 1920  
QY 1921 CILCOCYPTGSLSRVCDPEBQCPCKPGVIGQPCORCDMPFAEVTTNGCEVAVDSCPRAI 1980  
Db 1921 CILCOCYPTGSLSRVCDPEBQCPCKPGVIGQPCORCDMPFAEVTTNGCEVAVDSCPRAI 1980  
QY 1981 EAGIMWPRTRFGLPAAPCPKSGFSTAARHCDHRGMLPPLNFCTSLTFSELKGAERL 2040  
Db 1981 EAGIMWPRTRFGLPAAPCPKSGFSTAARHCDHRGMLPPLNFCTSLTFSELKGAERL 2040  
QY 2041 QNRESGLDSGRQOALALLRNATQHTAGYFGSDVAVAYQALATRLLAHESDORFGLSATQ 2100  
Db 2041 QNRESGLDSGRQOALALLRNATQHTAGYFGSDVAVAYQALATRLLAHESDORFGLSATQ 2100  
QY 2101 DVHFTENILRVGSALLDPANKRMHMLIOOTBEGTAMLOHYEAVASALAKONRHVTLSPF 2160  
Db 2101 DVHFTENILRVGSALLDPANKRMHMLIOOTBEGTAMLOHYEAVASALAKONRHVTLSPF 2160  
QY 2161 TIVTPTIVISVVRLDKGNFAGAKLPRYEALRGEOPPDETIVILPESVRETPPVVRPAG 2220  
Db 2161 TIVTPTIVISVVRLDKGNFAGAKLPRYEALRGEOPPDETIVILPESVRETPPVVRPAG 2220  
QY 2221 PGBAQBEBELARORRHPELSSGGAVASVYITRTLAGLLPHNYDDPKSLARPCKPIINT 2280  
Db 2221 PGBAQBEBELARORRHPELSSGGAVASVYITRTLAGLLPHNYDDPKSLARPCKPIINT 2280  
QY 2281 PIVYSISVHDEELPRALDKPVTVOFRLLETEBRTKPCIVFNHNSILVSGTGMARSAGE 2340  
Db 2281 PIVYSISVHDEELPRALDKPVTVOFRLLETEBRTKPCIVFNHNSILVSGTGMARSAGE 2340  
QY 2341 VVERNSHVSQCCNMHTSPAVLMDVSRRENGEILPLKLTLYVALGVTLAALLTFFFLTL 2400  
Db 2341 VVERNSHVSQCCNMHTSPAVLMDVSRRENGEILPLKLTLYVALGVTLAALLTFFFLTL 2400

QY 2401 LRIASNGHTRNLTAALGLAQLVFLGINDADIPACTVYAIILHFLYCTESWALLE 2460  
 Db 2401 LRIASNGHTRNLTAALGLAQLVFLGINDADIPACTVYAIILHFLYCTESWALLE 2460  
 QY 2461 ALHLRYALTEVDVNTGPMRFYFMGMGVPAPITGTAAGLDEGYNDFCMILSYDTLI 2520  
 Db 2461 ALHLRYALTEVDVNTGPMRFYFMGMGVPAPITGTAAGLDEGYNDFCMILSYDTLI 2520  
 QY 2521 WSPAPVPAFAVMSVFLYTLAARASCAAROGFEKKGPVSGIQSPFAVILLISATWLLAL 2580  
 Db 2521 WSPAPVPAFAVMSVFLYTLAARASCAAROGFEKKGPVSGIQSPFAVILLISATWLLAL 2580  
 QY 2581 LSVNSDTLFFHLYLFTKNCICIQPFIFLSVYVLSKVRALKACRKSPPDALTTKSTL 2640  
 Db 2581 LSVNSDTLFFHLYLFTKNCICIQPFIFLSVYVLSKVRALKACRKSPPDALTTKSTL 2640  
 QY 2641 TSSVNCSPYADGRYKOPYGDAGSLHSTRSGKQSPYIIFPLREESALNMGQGPGLG 2700  
 Db 2641 TSSVNCSPYADGRYKOPYGDAGSLHSTRSGKQSPYIIFPLREESALNMGQGPGLG 2700  
 QY 2701 DPGSLFLEGQDQHDPTDSDSLSLDDQSGSYASTHSDSEEBEEREAAPPGSG 2760  
 Db 2701 DPGSLFLEGQDQHDPTDSDSLSLDDQSGSYASTHSDSEEBEEREAAPPGSG 2760  
 QY 2761 WDSLIGPBAERLPLHSTPDGPGPKAPMPCDPCGTAKESGNAPEERLRENDALSR 2820  
 Db 2761 WDSLIGPBAERLPLHSTPDGPGPKAPMPCDPCGTAKESGNAPEERLRENDALSR 2820  
 QY 2821 EGSILGFLGSSAOPKIKKCLPTISEKSLPLPEOCTGSSRGSASBGRSGPP 2880  
 Db 2821 EGSILGFLGSSAOPKIKKCLPTISEKSLPLPEOCTGSSRGSASBGRSGPP 2880  
 QY 2881 RPPRQSLQEBQNGVPIAMSIKACTVDESSGSEFLFNFLLH 2923  
 Db 2881 RPPRQSLQEBQNGVPIAMSIKACTVDESSGSEFLFNFLLH 2923

## RESULT 2

US-09-916-849a-3  
 ; Sequence 3, Application US/09916849a  
 ; Publication No. US20030086934A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer, et al.  
 ; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents  
 ; TITLE OF INVENTION: Uses thereof  
 ; FILE REFERENCE: 2002850-0024  
 ; CURRENT APPLICATION NUMBER: US/09/916,849a  
 ; CURRENT FILING DATE: 2001-07-26  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Cadherin EGF  
 ; OTHER INFORMATION: LAG Seven Pass G-Type Receptor 2  
 US-09-916-849a-3

Query Match 100.0%; Score 15545; DB 11; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRSPATGVPLPTPEPPLLLLLLLLPPLGLDQVPCRSLSGRGSSGACAPMGWLCPS 60  
 Db 1 MRSPATGVPLPTPEPPLLLLLLLLPPLGLDQVPCRSLSGRGSSGACAPMGWLCPS 60  
 QY 61 SASNMLYTSRCDGTELTGHLVPHHNGLRWCESBAHIPLPAPPGCTWSCLLATIG 120  
 Db 61 SASNMLYTSRCDGTELTGHLVPHHNGLRWCESBAHIPLPAPPGCTWSCLLATIG 120  
 QY 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRRRKNVNTAPQ 180  
 Db 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRRRKNVNTAPQ 180

Db 121 GHLSPOGKLTLPPEHPCLKAPRLRCOSCKLAQAPGLRAGERSPEESLGRRRKNVNTAPQ 180  
 QY 181 FQPSYQATVBNOPAGTPVVASLRAIDPDBEGAGRLLETMDALPDSRNSQFSLDPYGA 240  
 Db 181 FQPSYQATVBNOPAGTPVVASLRAIDPDBEGAGRLLETMDALPDSRNSQFSLDPYGA 240  
 QY 241 VTTAEELDRETKSTHVRVTAQDHGMRRSALATLTLLVTDNDHDPVFEQOEKESLRE 300  
 Db 241 VTTAEELDRETKSTHVRVTAQDHGMRRSALATLTLLVTDNDHDPVFEQOEKESLRE 300  
 QY 301 NLEFYGVLVTRADGAPAPNANTLYRLRSGSSGSEBEVPIIDRSQVITRRGVDBREV 360  
 Db 301 NLEFYGVLVTRADGAPAPNANTLYRLRSGSSGSEBEVPIIDRSQVITRRGVDBREV 360  
 QY 361 ESYOLTVEASDOGRDPPRSTTAAVFVSVEDDNNAAQFSEKRYVYVREDVTGAPVLR 420  
 Db 361 ESYOLTVEASDOGRDPPRSTTAAVFVSVEDDNNAAQFSEKRYVYVREDVTGAPVLR 420  
 QY 421 VTTASDRKGSNAVHVSIMSGNARGQFYLDAGTALDVSPLDYETTKETTLRVRAQDG 480  
 Db 421 VTTASDRKGSNAVHVSIMSGNARGQFYLDAGTALDVSPLDYETTKETTLRVRAQDG 480  
 QY 481 RPLSNVSGLVTVQVLIJNNAPIFVSTPFOATVLSVPIGYLVIAHOAIDADGDNARL 540  
 Db 481 RPLSNVSGLVTVQVLIJNNAPIFVSTPFOATVLSVPIGYLVIAHOAIDADGDNARL 540  
 QY 541 EYRLAGVGHDPFTTINNCGMISVAELDBREVDYFSGYEARDHGTALTAASVSVTV 600  
 Db 541 EYRLAGVGHDPFTTINNCGMISVAELDBREVDYFSGYEARDHGTALTAASVSVTV 600  
 QY 601 LDVNDNPPFTTQPEYTVRLMEDAAVGTSVTVSANDBAHSVITYQITSGTNRNRSITS 660  
 Db 601 LDVNDNPPFTTQPEYTVRLMEDAAVGTSVTVSANDBAHSVITYQITSGTNRNRSITS 660  
 QY 661 OSGGGLVSLALPLDYKLEROYLAVTASDGTROTAQIVNVNDANTRHPFQSSHTVN 720  
 Db 661 OSGGGLVSLALPLDYKLEROYLAVTASDGTROTAQIVNVNDANTRHPFQSSHTVN 720  
 QY 721 VNEBRPAGTTVVLISATDEDTGENARITTFEMEDSIPOFRIDADGAVTTOAELEDEDOVS 780  
 Db 721 VNEBRPAGTTVVLISATDEDTGENARITTFEMEDSIPOFRIDADGAVTTOAELEDEDOVS 780  
 QY 781 YTLAITARDNGIPQKSTTYTLELIVNDVNDNAQPLFDSIQSGSYEDVPPTSTVLQISAT 840  
 Db 781 YTLAITARDNGIPQKSTTYTLELIVNDVNDNAQPLFDSIQSGSYEDVPPTSTVLQISAT 840  
 QY 841 DRDSGLNGRVFTYFQGGDDGDDGDFIVESTIGIVKTLRLREBNVAQVILRAYAVDKGMP 900  
 Db 841 DRDSGLNGRVFTYFQGGDDGDDGDFIVESTIGIVKTLRLREBNVAQVILRAYAVDKGMP 900  
 QY 901 ARTPEVTVTVLVNDNPPVFEQDEPDVFEENSPIGLAAVARVATPDDEGTNAQIMYQI 960  
 Db 901 ARTPEVTVTVLVNDNPPVFEQDEPDVFEENSPIGLAAVARVATPDDEGTNAQIMYQI 960  
 QY 961 VEGNIPPEVFOLDIFSGELTALVDLDYEDREPEYLVIOATSAPLVSRAVTHVRLDRNDP 1020  
 Db 961 VEGNIPPEVFOLDIFSGELTALVDLDYEDREPEYLVIOATSAPLVSRAVTHVRLDRNDP 1020  
 QY 1021 PVYGNPEILFNNYVYNNSSSFGGATGRVAPHPDLSLTSYPERENELSIYVLNASTG 1080  
 Db 1021 PVYGNPEILFNNYVYNNSSSFGGATGRVAPHPDLSLTSYPERENELSIYVLNASTG 1080  
 QY 1081 ELKLSRALDNNRPLLEAIMSVLVSQGVASVTAQCALRTYITITDEMLTSTILRLDEMSPER 1140  
 Db 1081 ELKLSRALDNNRPLLEAIMSVLVSQGVASVTAQCALRTYITITDEMLTSTILRLDEMSPER 1140  
 QY 1141 FLSPILGLFTQAVNAATATATPPDHVVVNVNQRDTPAGGHLIANTSLSGQPPGSGGPPFL 1200  
 Db 1141 FLSPILGLFTQAVNAATATATPPDHVVVNVNQRDTPAGGHLIANTSLSGQPPGSGGPPFL 1200  
 QY 1201 PSEDLQERLYNRSILTAISAQVLPFDNDCIAREPCENTMRCSVLRPFSSAPFIASSS 1260  
 Db 1201 PSEDLQERLYNRSILTAISAQVLPFDNDCIAREPCENTMRCSVLRPFSSAPFIASSS 1260

QY 1261 VLEFPIHPVGLNCRCPGFTGDCYCEYVDLCTSRPCGPHGRCSRBSGGYTCICRDGYG 1320  
 DB 1261 VLEFPIHPVGLNCRCPGFTGDCYCEYVDLCTSRPCGPHGRCSRBSGGYTCICRDGYG 1320  
 QY 1321 EHCVEVARSGRCTPGVCKNGCTCNLLVGGFKCDPSGDEKPYCOVTRSPASFTIP 1380  
 DB 1321 EHCVEVARSGRCTPGVCKNGCTCNLLVGGFKCDPSGDEKPYCOVTRSPASFTIP 1380  
 QY 1381 RGLRSHFHTLSPATKRDGLLLNGRPNKHDPALEVOEVOULTPSASESTTYS 1440  
 DB 1381 RGLRSHFHTLSPATKRDGLLLNGRPNKHDPALEVOEVOULTPSASESTTYS 1440  
 QY 1441 PFWPGVSDGQMTVOLKTYNKEPLGQGTLPQSPSEOKVAVVWVDCDITVALRFSVVG 1500  
 DB 1441 PFWPGVSDGQMTVOLKTYNKEPLGQGTLPQSPSEOKVAVVWVDCDITVALRFSVVG 1500  
 QY 1501 NYSCAAQGTQSGSKSLDTLGPILLGVPPLPESFVRMRQFVGCNRNLOVDSRHIDMAD 1560  
 DB 1501 NYSCAAQGTQSGSKSLDTLGPILLGVPPLPESFVRMRQFVGCNRNLOVDSRHIDMAD 1560  
 QY 1561 FIANNGTVPCCPAKQVCDNTCHNGGTCVNDADAFSCBCLFPGGKSCAQBMANPOHFL 1620  
 DB 1561 FIANNGTVPCCPAKQVCDNTCHNGGTCVNDADAFSCBCLFPGGKSCAQBMANPOHFL 1620  
 QY 1621 GSSLVAMHGLSLPISQPWYLSLMTFRTOADGVLLQAITRGRSTITLQLEHGHVMSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISQPWYLSLMTFRTOADGVLLQAITRGRSTITLQLEHGHVMSVEGT 1680  
 QY 1681 GLQASSLRLEPGANGDHHAQALGASGPGHALLSPYGOORABGNIGPLHGLHLS 1740  
 DB 1681 GLQASSLRLEPGANGDHHAQALGASGPGHALLSPYGOORABGNIGPLHGLHLS 1740  
 QY 1741 NITVGGIPBAGVARGFRGCTQGVNVDTPREGVNSLDPSHGSINVBQCSLPDCCSN 1800  
 DB 1741 NITVGGIPBAGVARGFRGCTQGVNVDTPREGVNSLDPSHGSINVBQCSLPDCCSN 1800  
 QY 1801 PCPANSYCSNDMSYSQSCDPGYGDNCTNVCDLNPCEHQSVCTRKPSAHGYTCCEPPN 1860  
 DB 1801 PCPANSYCSNDMSYSQSCDPGYGDNCTNVCDLNPCEHQSVCTRKPSAHGYTCCEPPN 1860  
 QY 1861 YLGPYCETRIDOCPRGWMHPTCGPCNCDVSGPBDCKNTSGECHKCNHRRPSPPT 1920  
 DB 1861 YLGPYCETRIDOCPRGWMHPTCGPCNCDVSGPBDCKNTSGECHKCNHRRPSPPT 1920  
 QY 1921 CLICDCYPTGSLRVCDBEDGQCPCKRGVIGROCDRCNPFATVINGCEVNTDSCPRAT 1980  
 DB 1921 CLICDCYPTGSLRVCDBEDGQCPCKRGVIGROCDRCNPFATVINGCEVNTDSCPRAT 1980  
 QY 1981 BAGIWPRTFGLPAAAPCPKSGFGTAVRHCDHRGMLPMLFNCTISITFSELKGFABRL 2040  
 DB 1981 BAGIWPRTFGLPAAAPCPKSGFGTAVRHCDHRGMLPMLFNCTISITFSELKGFABRL 2040  
 QY 2041 QRBESGLDSGRSOQLALLNATONHTAGVSGDVKAYAYOLATLHSHESQORFGLSAQ 2100  
 DB 2041 QRBESGLDSGRSOQLALLNATONHTAGVSGDVKAYAYOLATLHSHESQORFGLSAQ 2100  
 QY 2101 DVAFTEMLRVSALDPTANKRMWELIQTEGTAMLQHYEAYASALQONMHTYLSPP 2160  
 DB 2101 DVAFTEMLRVSALDPTANKRMWELIQTEGTAMLQHYEAYASALQONMHTYLSPP 2160  
 QY 2161 TIYTPNIVISVRLDKNPAAGAKLPRYEALRGEPDLETTVILPESVFEETPPVVRPAG 2220  
 DB 2161 TIYTPNIVISVRLDKNPAAGAKLPRYEALRGEPDLETTVILPESVFEETPPVVRPAG 2220  
 QY 2221 PGEAOBEEELARORRHELSOGEAVASVIYRTLAGLPHNYDPOGRSLRVKRPINT 2280  
 DB 2221 PGEAOBEEELARORRHELSOGEAVASVIYRTLAGLPHNYDPOGRSLRVKRPINT 2280  
 QY 2281 PVSISVHDEBELLPRALDKPVTVQFRLLEBETKPICVFMNHSILVSGTGMASARGE 2340  
 DB 2281 PVSISVHDEBELLPRALDKPVTVQFRLLEBETKPICVFMNHSILVSGTGMASARGE 2340

QY 2341 VFRNESHVSQCNMHTSPAVLMDVSRRENGEILPLKTLTVVALGVTLAALLTEFFLTL 2400  
 DB 2341 VFRNESHVSQCNMHTSPAVLMDVSRRENGEILPLKTLTVVALGVTLAALLTEFFLTL 2400  
 QY 2401 LRIILSNQGRIRNLTALGLAQVFLGINDADLPFACTVYAILHFLYLTCEFWALLIE 2460  
 DB 2401 LRIILSNQGRIRNLTALGLAQVFLGINDADLPFACTVYAILHFLYLTCEFWALLIE 2460  
 QY 2461 ALHLRYALTEVADVNTGPMRFVYMLGMGVPAFTGLAAGLDDEGYNPFCMLSYDTLI 2520  
 DB 2461 ALHLRYALTEVADVNTGPMRFVYMLGMGVPAFTGLAAGLDDEGYNPFCMLSYDTLI 2520  
 QY 2521 WSPAGVPAVAVMSVFLYTLAARASCAQORQFEKGPVSGLOPSFAVLLLSATWLLAL 2580  
 DB 2521 WSPAGVPAVAVMSVFLYTLAARASCAQORQFEKGPVSGLOPSFAVLLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLPAFTNCIOGPPIFLSYVYLSKVRKALKACGRKSPDALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLPAFTNCIOGPPIFLSYVYLSKVRKALKACGRKSPDALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLYOPYGDAGSLHSTSRGSKSQSYIPLAREBSALNPGQPPGLG 2700  
 DB 2641 TSSYNCPSPYADGRLYOPYGDAGSLHSTSRGSKSQSYIPLAREBSALNPGQPPGLG 2700  
 QY 2701 DFGSLFLRQDQDHPDITDSDSLSEDDQSGSVASTHSDBSEEBEERBAEPGEG 2760  
 DB 2701 DFGSLFLRQDQDHPDITDSDSLSEDDQSGSVASTHSDBSEEBEERBAEPGEG 2760  
 QY 2761 WDSLIGPABRLPLHSTRDGPGPKAPWPDFTTAKESGNAPEERLRENDALSR 2820  
 DB 2761 WDSLIGPABRLPLHSTRDGPGPKAPWPDFTTAKESGNAPEERLRENDALSR 2820  
 QY 2821 EGSIGPLPSSAQPHKGLIKKKCLPTISEKSLILPLEQCTGSSRGSASBSGCGPP 2880  
 DB 2821 EGSIGPLPSSAQPHKGLIKKKCLPTISEKSLILPLEQCTGSSRGSASBSGCGPP 2880  
 QY 2881 RPPRQSIQEQINGVPIAMSIKAGTVDBSSGSFLLPFFNLH 2923  
 DB 2881 RPPRQSIQEQINGVPIAMSIKAGTVDBSSGSFLLPFFNLH 2923

RESULT 3  
 US-10-174-677-29  
 ; Sequence 29, Application US/10174677  
 ; Publication No. US20030190704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xie, Ting  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR  
 ; FILE REFERENCE: 40716(1P-012)  
 ; CURRENT APPLICATION NUMBER: US/10/174.677  
 ; CURRENT FILING DATE: 2002-06-19  
 ; NUMBER OF SEQ ID NOS: 117  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 29  
 ; LENGTH: 2923  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-174-677-29

Query Match 100.0%; Score 15545; DB 12; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPATGVPLPPEPPLLLLLLLPPLLDQVGPCHSLGSRGSSGACAPMGLCP 60  
 DB 1 MSPATGVPLPPEPPLLLLLLLPPLLDQVGPCHSLGSRGSSGACAPMGLCP 60  
 QY 61 SASNMLTYSRCKRDGTELTGHLVPHHGLRWYCESEAHITLPAPBECFMSCLLGIG 120  
 DB 61 SASNMLTYSRCKRDGTELTGHLVPHHGLRWYCESEAHITLPAPBECFMSCLLGIG 120  
 QY 121 GHLSPQSKLTLEEARPCAKAPLRSCSCGLAOPGLRAGERSPEBSLGRGRANNTAPO 180  
 DB 121 GHLSPQSKLTLEEARPCAKAPLRSCSCGLAOPGLRAGERSPEBSLGRGRANNTAPO 180

Db 121 GHLSPQKLTLPBEPCLKAPRLCQSCKLAQAPGLRAGERSPEESLGRRRRNVNTPAQ 180  
 Qy 181 FQPSYQATVPENQAGTPVVASLRAIDPDEGEAGLEXTMALFDRSNOQFSLDPTVGA 240  
 Db 181 FQPSYQATVPENQAGTPVVASLRAIDPDEGEAGLEXTMALFDRSNOQFSLDPTVGA 240  
 Qy 241 VTLAEELREKTSVFRVTAQDHGMPRRSALATITLVTDTNDHPVFEQOYKESLRE 300  
 Db 241 VTLAEELREKTSVFRVTAQDHGMPRRSALATITLVTDTNDHPVFEQOYKESLRE 300  
 Qy 301 NLEVEYEVLTARAATGDAPPNANILYRLLEGSGSPSEVEFEIDPESGVIRTRGPVDRREV 360  
 Db 301 NLEVEYEVLTARAATGDAPPNANILYRLLEGSGSPSEVEFEIDPESGVIRTRGPVDRREV 360  
 Qy 361 ESYQVLTASDQGRDPGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREBDVTPCAPYLK 420  
 Db 361 ESYQVLTASDQGRDPGRSTTAAVFLSVEDNDNAPQFSEKRYVQVREBDVTPCAPYLK 420  
 Qy 421 VTAASDRDGSNAVHYHSIMSGNARQFTLDAQTGALDVVSPLDYETTKETLRVAQDGG 480  
 Db 421 VTAASDRDGSNAVHYHSIMSGNARQFTLDAQTGALDVVSPLDYETTKETLRVAQDGG 480  
 Qy 481 RPLPSNVSGLVTVOYLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADGDNRL 540  
 Db 481 RPLPSNVSGLVTVOYLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALDADGDNRL 540  
 Qy 541 EYRLAGVGHDPPTINNCGWISVAABLDREVDVFSFGVEARDHGTALPASASVTV 600  
 Db 541 EYRLAGVGHDPPTINNCGWISVAABLDREVDVFSFGVEARDHGTALPASASVTV 600  
 Qy 601 LDVNDNNTFPOPEYTVLNEBDAVAGTSVTVSANDRAHSYITQTISGNTRRFSTTS 660  
 Db 601 LDVNDNNTFPOPEYTVLNEBDAVAGTSVTVSANDRAHSYITQTISGNTRRFSTTS 660  
 Qy 661 QSGGGLVSLALPLDYKLEROYLAATASDGTRODPAQIVANVTANTHRPVOSSHVTYN 720  
 Db 661 QSGGGLVSLALPLDYKLEROYLAATASDGTRODPAQIVANVTANTHRPVOSSHVTYN 720  
 Qy 721 VNEBRPAGTTVVLISATDEDTGENARITYFMEBSIPQFRIDADTCAVTTQAELEDYEDVS 780  
 Db 721 VNEBRPAGTTVVLISATDEDTGENARITYFMEBSIPQFRIDADTCAVTTQAELEDYEDVS 780  
 Qy 781 YTLAITADNGIPQKSDPTTYLEIIVNDVNDNAPQFLRBSYQGSYVEDVPPPTSVLQISAT 840  
 Db 781 YTLAITADNGIPQKSDPTTYLEIIVNDVNDNAPQFLRBSYQGSYVEDVPPPTSVLQISAT 840  
 Qy 841 DRDSGLNRRVFTYFQGGDGDGDFIVESTSGIVRTLRDLRENVAAQVLRAYAVDKGMP 900  
 Db 841 DRDSGLNRRVFTYFQGGDGDGDFIVESTSGIVRTLRDLRENVAAQVLRAYAVDKGMP 900  
 Qy 901 ARTPEVTVTVLDVNDNPPVFEODEFDVFEENSPITGLAVARVYATDDEBGTNAQIMYQI 960  
 Db 901 ARTPEVTVTVLDVNDNPPVFEODEFDVFEENSPITGLAVARVYATDDEBGTNAQIMYQI 960  
 Qy 961 VEGNIPBEVPOLDISGELTALVDLDYEDRPREVUVIQTASAPLYSRATVYHRLDNRNP 1020  
 Db 961 VEGNIPBEVPOLDISGELTALVDLDYEDRPREVUVIQTASAPLYSRATVYHRLDNRNP 1020  
 Qy 1021 PVLGNFELLFNNYVTNRSSFPGCAIGRVPADPDIPSLTYSFPRGNELSIYVLLNASTG 1080  
 Db 1021 PVLGNFELLFNNYVTNRSSFPGCAIGRVPADPDIPSLTYSFPRGNELSIYVLLNASTG 1080  
 Qy 1081 ELKLSRALDNNRPLEAIVSVSDGVHSVTAQCALRTVITIDEMLTHTSITLRLDMSPER 1140  
 Db 1081 ELKLSRALDNNRPLEAIVSVSDGVHSVTAQCALRTVITIDEMLTHTSITLRLDMSPER 1140  
 Qy 1141 FLSPILGLFIOAVAAATLATPDDHVVVENVORDTAPGHHIINVSLSVGQPPGPGGPPFL 1200  
 Db 1141 FLSPILGLFIOAVAAATLATPDDHVVVENVORDTAPGHHIINVSLSVGQPPGPGGPPFL 1200  
 Qy 1201 PSEDLQERLYLNRSILTAISAQVLPFDNITCLREPCENYKRCVSLRFDSSAPFIASS 1260  
 Db 1201 PSEDLQERLYLNRSILTAISAQVLPFDNITCLREPCENYKRCVSLRFDSSAPFIASS 1260

Qy 1261 VLFRIHPVGLRCRCPGFTGADYCEITEVDLCYRPPCGPHRCRSRREGYTCLCRDGYTG 1320  
 Db 1261 VLFRIHPVGLRCRCPGFTGADYCEITEVDLCYRPPCGPHRCRSRREGYTCLCRDGYTG 1320  
 Qy 1321 EHCESVARSGRCTPGVCNKGCTCVNLLVGFKCDPCEGDFEKPVCQVTTSPHSPITF 1380  
 Db 1321 EHCESVARSGRCTPGVCNKGCTCVNLLVGFKCDPCEGDFEKPVCQVTTSPHSPITF 1380  
 Qy 1381 RGLRORHFTLALSPATKBDGLLXNGRFPEKHDPALEVIQOQVOLTSPAGSSTTVS 1440  
 Db 1381 RGLRORHFTLALSPATKBDGLLXNGRFPEKHDPALEVIQOQVOLTSPAGSSTTVS 1440  
 Qy 1441 PFVPGVSDGQMTVOLKXYNKPILLQGTGLPOGSEOKVAVVTYDGDCTGVALRFGSVLG 1500  
 Db 1441 PFVPGVSDGQMTVOLKXYNKPILLQGTGLPOGSEOKVAVVTYDGDCTGVALRFGSVLG 1500  
 Qy 1501 NYSCAQTGQGSKKSIDLTPILLGGVPDLPSFPVRMQFVGCNENLOVDSHHIMAD 1560  
 Db 1501 NYSCAQTGQGSKKSIDLTPILLGGVPDLPSFPVRMQFVGCNENLOVDSHHIMAD 1560  
 Qy 1561 FIANNGTVPGCCPAKKNVCDNNTCHNGGTGVNOMDAPSCBCEPLGFGKSCAQEMANPOHFL 1620  
 Db 1561 FIANNGTVPGCCPAKKNVCDNNTCHNGGTGVNOMDAPSCBCEPLGFGKSCAQEMANPOHFL 1620  
 Qy 1621 GSSLVAMHGLSLPISQPMYLSLMEFTRQADGVLLQAITRGHSTITTLQLRGHVMSVEGT 1680  
 Db 1621 GSSLVAMHGLSLPISQPMYLSLMEFTRQADGVLLQAITRGHSTITTLQLRGHVMSVEGT 1680  
 Qy 1681 GLQASSRLRBPGRANDGMHQAOLALGASGPGHAIISPTYGQORAGNIGPRLHGLHS 1740  
 Db 1681 GLQASSRLRBPGRANDGMHQAOLALGASGPGHAIISPTYGQORAGNIGPRLHGLHS 1740  
 Qy 1741 NITVGGIIPGAGVARGFRGCLQGVRSPTBEGNSLDPBSHGSINVEQGSJLDPDCSN 1800  
 Db 1741 NITVGGIIPGAGVARGFRGCLQGVRSPTBEGNSLDPBSHGSINVEQGSJLDPDCSN 1800  
 Qy 1801 PCPANSYCSNDMDYSYSCDPGYGDMCTVNCDLNPECNHSVCTRKSPHGYTCECPN 1860  
 Db 1801 PCPANSYCSNDMDYSYSCDPGYGDMCTVNCDLNPECNHSVCTRKSPHGYTCECPN 1860  
 Qy 1861 YLGPYCESTRIDOPCPRMWHPHTCGPCNCDVSKGFPDCKNTSGECHKENHPRPSSPT 1920  
 Db 1861 YLGPYCESTRIDOPCPRMWHPHTCGPCNCDVSKGFPDCKNTSGECHKENHPRPSSPT 1920  
 Qy 1921 CLLCDYPTGSLSRVCPEDGQCPCKPGVIGROCDRCNDFPAEYTTNGCEVYNSCPRAI 1980  
 Db 1921 CLLCDYPTGSLSRVCPEDGQCPCKPGVIGROCDRCNDFPAEYTTNGCEVYNSCPRAI 1980  
 Qy 1981 EAGIWPRTFGLPAAAPCEKSGFGTAVRHCDERHGLPNNLFNCTSIITSELSKGAERL 2040  
 Db 1981 EAGIWPRTFGLPAAAPCEKSGFGTAVRHCDERHGLPNNLFNCTSIITSELSKGAERL 2040  
 Qy 2041 QRNESGLDSRQOALLLNATQHTAGYGSVYKAYVQIATLILAHESYORPGLSATQ 2100  
 Db 2041 QRNESGLDSRQOALLLNATQHTAGYGSVYKAYVQIATLILAHESYORPGLSATQ 2100  
 Qy 2101 DVHFTBMLAVGSALDPTANKRMELIQQTEGGTAMLLQYBYAASALQNMHTYLSPF 2160  
 Db 2101 DVHFTBMLAVGSALDPTANKRMELIQQTEGGTAMLLQYBYAASALQNMHTYLSPF 2160  
 Qy 2161 TIVTPNIVISVRLDKGNFAGATLPRYBALRGEOPDLETTVILPESVRETPPVNAPAG 2220  
 Db 2161 TIVTPNIVISVRLDKGNFAGATLPRYBALRGEOPDLETTVILPESVRETPPVNAPAG 2220  
 Qy 2221 PGEAOEPEELARBORHPELSQGEAAVASVITYTLAGLPHANDPDKRSIRVPRPIINT 2280  
 Db 2221 PGEAOEPEELARBORHPELSQGEAAVASVITYTLAGLPHANDPDKRSIRVPRPIINT 2280  
 Qy 2281 PVSISVHDBEELPRLADKPVTVQFRLLETBERTRPICVFNNHSILVSGTGMARSARGE 2340  
 Db 2281 PVSISVHDBEELPRLADKPVTVQFRLLETBERTRPICVFNNHSILVSGTGMARSARGE 2340

```

QY 2341 VVFRNSHVSQCQNMNTSPAVIDMVSRRNGEILPLKTLTYVALGVTLAALLTFFELTL 2400
DB 2341 VVFRNSHVSQCQNMNTSPAVIDMVSRRNGEILPLKTLTYVALGVTLAALLTFFELTL 2400
QY 2401 LRLRNSHGIRNLTALGLAQVFLGIGNADLPACTVLAALLHPTLCFFSALLL 2460
DB 2401 LRLRNSHGIRNLTALGLAQVFLGIGNADLPACTVLAALLHPTLCFFSALLL 2460
QY 2461 ALHLRYALTEVRDVNTGPMRFYTMGMGVPAPFTGLAVGLDPGCGNPPFCMLSIYDTLI 2520
DB 2461 ALHLRYALTEVRDVNTGPMRFYTMGMGVPAPFTGLAVGLDPGCGNPPFCMLSIYDTLI 2520
QY 2521 WSPAGVAVAVSVSVFLYIILAAASCAAQOGFEGKGVSGLOPSAVILLLSATWLLAL 2580
DB 2521 WSPAGVAVAVSVSVFLYIILAAASCAAQOGFEGKGVSGLOPSAVILLLSATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFATCNCCIQQPFILSVVLSKEVRKALKACSRKPSDPDLATTKSTL 2640
DB 2581 LSVNSDTLLFHYLFATCNCCIQQPFILSVVLSKEVRKALKACSRKPSDPDLATTKSTL 2640
QY 2641 TSSYNCPSPYADGRLYQPYDSAGSLHSTSRSGSKQPSYIPLLRRESALNPQCGPGLG 2700
DB 2641 TSSYNCPSPYADGRLYQPYDSAGSLHSTSRSGSKQPSYIPLLRRESALNPQCGPGLG 2700
QY 2701 DPGLFLBEGDDQHDPTDSDSLBDQSGSYASTHSDSEEEEEEBAAPFEGG 2760
DB 2701 DPGLFLBEGDDQHDPTDSDSLBDQSGSYASTHSDSEEEEEEBAAPFEGG 2760
QY 2761 WBSLLPBGARLPLHSTPKXGPGPGKAPWPGRGTTAKSSGNGAPBEERLNRNGALR 2820
DB 2761 WBSLLPBGARLPLHSTPKXGPGPGKAPWPGRGTTAKSSGNGAPBEERLNRNGALR 2820
QY 2821 BGSIGLPGSSAOPHKGILLKKCLPTISEKSLRLPLBQCTSSSGSSASBSRGSGPP 2880
DB 2821 BGSIGLPGSSAOPHKGILLKKCLPTISEKSLRLPLBQCTSSSGSSASBSRGSGPP 2880
QY 2881 RPPRQSLQBLQNGVMPIMANSIKAGTYDEBSGSEELFFNFPLH 2923
DB 2881 RPPRQSLQBLQNGVMPIMANSIKAGTYDEBSGSEELFFNFPLH 2923

```

```

RESULT 4
US-10-120-801-53
; Sequence 53, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkete, Richard
; APPLICANT: Padigattu, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehriban, Ruad
; APPLICANT: Topper, James N
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120, 801
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241

```

```

; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 53
; LENGTH: 2923
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-120-801-53

```

```

Query Match 100.0%; Score 15545; DB 12; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MRSPTGYPLPPPPPLLLLLLLLLLPPPLDQVPCPSLSGRSGSGACAPMCMCLPS 60
DB 1 MRSPTGYPLPPPPPLLLLLLLLLLPPPLDQVPCPSLSGRSGSGACAPMCMCLPS 60
QY 61 SASNMLYTSRCRDAGTELTHLVPHDGLRVYCPSESAHPLPPAPGCPMSCGLGIG 120
DB 61 SASNMLYTSRCRDAGTELTHLVPHDGLRVYCPSESAHPLPPAPGCPMSCGLGIG 120
QY 121 GHLSPQKLTLPBEPHCILAPRLRCQSCKLAQAPGLARABSPBSLGGRRKRVNTAPQ 180
DB 121 GHLSPQKLTLPBEPHCILAPRLRCQSCKLAQAPGLARABSPBSLGGRRKRVNTAPQ 180
QY 181 FQPPSYQATVPENQAGTPVASTRAIDPDEBAGLFTTMALPBRNPFSLDPTGA 240
DB 181 FQPPSYQATVPENQAGTPVASTRAIDPDEBAGLFTTMALPBRNPFSLDPTGA 240
QY 241 VTTAEELREKSTVFVFTADHGMPPRSALATLTITVTDTNDHPVEQOEYESLRE 300
DB 241 VTTAEELREKSTVFVFTADHGMPPRSALATLTITVTDTNDHPVEQOEYESLRE 300
QY 301 NLEVGVEVLTVATNDGAPPNANILYRLBSGSGSPSEVFIDPESGYIRRGVDRREV 360
DB 301 NLEVGVEVLTVATNDGAPPNANILYRLBSGSGSPSEVFIDPESGYIRRGVDRREV 360
QY 361 ESYQULTVEASDQGRPGPSTTAAVFLSVEDNDNDAPQSEKRYVVOVEDVTGAPVLR 420
DB 361 ESYQULTVEASDQGRPGPSTTAAVFLSVEDNDNDAPQSEKRYVVOVEDVTGAPVLR 420
QY 421 VTASDRDKSNAVHYHSIMSGNARQFYLDAGTGLDVVSPLDVETTYETLRVAAODGG 480
DB 421 VTASDRDKSNAVHYHSIMSGNARQFYLDAGTGLDVVSPLDVETTYETLRVAAODGG 480
QY 481 RPPLSNVSGLVTVQVLDINDNAPITVSTPQATVLESYPLGYLVHQAIDADAGDNRL 540
DB 481 RPPLSNVSGLVTVQVLDINDNAPITVSTPQATVLESYPLGYLVHQAIDADAGDNRL 540
QY 541 EYRLAGVGHDPFTTNNNGWISVAEILDBREVDSYSGVBARHGTPLTASASVTV 600
DB 541 EYRLAGVGHDPFTTNNNGWISVAEILDBREVDSYSGVBARHGTPLTASASVTV 600
QY 601 LDVNDNNPFTQPEXTYVRLNEDAAVGTSVYSAVDRDASHYITQITSGNTRNFSITS 660
DB 601 LDVNDNNPFTQPEXTYVRLNEDAAVGTSVYSAVDRDASHYITQITSGNTRNFSITS 660
QY 661 QSGGGIVSLAPLDYKLERQYTLATASDGTRODPAQIVNNTDANTRPVQSSHYYN 720
DB 661 QSGGGIVSLAPLDYKLERQYTLATASDGTRODPAQIVNNTDANTRPVQSSHYYN 720
QY 721 VNEDRPAGTIVVLSATBEDTGENARITYFMEDSIPQRIADJGAVTTQALDEDQVS 780
DB 721 VNEDRPAGTIVVLSATBEDTGENARITYFMEDSIPQRIADJGAVTTQALDEDQVS 780
QY 781 YTLAITARDNGIPQSDTYYEILVNDVNDNAPQFLRDSYQGSYVEDVPFTSVLQISAT 840
DB 781 YTLAITARDNGIPQSDTYYEILVNDVNDNAPQFLRDSYQGSYVEDVPFTSVLQISAT 840

```



```

QY 841 DEDSGINGEFTTPOGDDGDDGFIVESTSGIVTRLRRLRENNVAQVYLAAVADKAMP 900
DB 841 DEDSGINGEFTTPOGDDGDDGFIVESTSGIVTRLRRLRENNVAQVYLAAVADKAMP 900
QY 901 ATTPMEVTVTVLDVNDNPFVEFODEPDVYEEKSPGLAVARVATDPDEBGTNAQIMYOI 960
DB 901 ATTPMEVTVTVLDVNDNPFVEFODEPDVYEEKSPGLAVARVATDPDEBGTNAQIMYOI 960
QY 961 VEGNIEVEFOLDIFSGELTALVDLDEDEPEYLVVQATSAPLVSRATYHVLBDNDMP 1020
DB 961 VEGNIEVEFOLDIFSGELTALVDLDEDEPEYLVVQATSAPLVSRATYHVLBDNDMP 1020
QY 1021 PVLGNFEILFNNVYTNRSSPPGAGIARVAPNDPDISLTYSPERGENELSLVLTNASTG 1080
DB 1021 PVLGNFEILFNNVYTNRSSPPGAGIARVAPNDPDISLTYSPERGENELSLVLTNASTG 1080
QY 1081 ELKSRALDNNRPLBAINSVUSDSGASTTAQCALAVTITDBMLTHSITTLAEDMSPER 1140
DB 1081 ELKSRALDNNRPLBAINSVUSDSGASTTAQCALAVTITDBMLTHSITTLAEDMSPER 1140
QY 1141 FLSPLILGLEIOAVATLTPPDHVVENVQORDTAPAGHILNVSLSVQPPGPGGPPPL 1200
DB 1141 FLSPLILGLEIOAVATLTPPDHVVENVQORDTAPAGHILNVSLSVQPPGPGGPPPL 1200
QY 1201 PSEDIQERLYNRSLSITLTAISAQVLPEDNITCLREPCENYMRCSVLPDSSAPFIASS 1260
DB 1201 PSEDIQERLYNRSLSITLTAISAQVLPEDNITCLREPCENYMRCSVLPDSSAPFIASS 1260
QY 1261 VLFPRPHVGGIARCRPGFTDCTEVDLCYSRCGPHGRCSREGGYTCLCRDGYTG 1320
DB 1261 VLFPRPHVGGIARCRPGFTDCTEVDLCYSRCGPHGRCSREGGYTCLCRDGYTG 1320
QY 1321 EHCEVASABSGRCPTGVCNKGTCVNLVYGFKDCPSGDPEKPYCOVTRSPFAPSFTF 1380
DB 1321 EHCEVASABSGRCPTGVCNKGTCVNLVYGFKDCPSGDPEKPYCOVTRSPFAPSFTF 1380
QY 1381 RGLRORFHEFTLALSPATKERDGLLYNGRFNEKHDFVALEVIQEOVLTFSGESTTVS 1440
DB 1381 RGLRORFHEFTLALSPATKERDGLLYNGRFNEKHDFVALEVIQEOVLTFSGESTTVS 1440
QY 1441 PVPVGSVSGOMHTVQKXNNKPLIGOTGLPGSPBOKYAVVTVNGCDGVNLRGSGVIG 1500
DB 1441 PVPVGSVSGOMHTVQKXNNKPLIGOTGLPGSPBOKYAVVTVNGCDGVNLRGSGVIG 1500
QY 1501 NYSACAQGTGGSKXSLDTLGPBLLGVPDLPESPFVRARQFVGCNRLQVDSRIIDMAD 1560
DB 1501 NYSACAQGTGGSKXSLDTLGPBLLGVPDLPESPFVRARQFVGCNRLQVDSRIIDMAD 1560
QY 1561 FIANNGTVPGCAPKKNVCDNNTCHNGGTCVNOMDAFSCCEPLGFGGKSCAQMAMPQHFL 1620
DB 1561 FIANNGTVPGCAPKKNVCDNNTCHNGGTCVNOMDAFSCCEPLGFGGKSCAQMAMPQHFL 1620
QY 1621 GSSLVAMHGLSLPIISOPWYLSLMPFTRROADGVLLAITGGRSTITLQLEBGMVLSVBST 1680
DB 1621 GSSLVAMHGLSLPIISOPWYLSLMPFTRROADGVLLAITGGRSTITLQLEBGMVLSVBST 1680
QY 1681 GLQASSLRLEPRANDGDWHHAQLALGASGPGHAILSFYGOQBAEGNLGRLILGLHS 1740
DB 1681 GLQASSLRLEPRANDGDWHHAQLALGASGPGHAILSFYGOQBAEGNLGRLILGLHS 1740
QY 1741 NITVGGIPEPAGVARGFPGCLQGVRSVTPREGVNSLDPSEBSINVEOGCSLPDPCSN 1800
DB 1741 NITVGGIPEPAGVARGFPGCLQGVRSVTPREGVNSLDPSEBSINVEOGCSLPDPCSN 1800
QY 1801 PCPANSYCSNDMSYSCSDPGYGDNCNVCDLNPCEHOSVCTRKPSPHGYTORCPN 1860
DB 1801 PCPANSYCSNDMSYSCSDPGYGDNCNVCDLNPCEHOSVCTRKPSPHGYTORCPN 1860
QY 1861 YLGPHYCETRIDQPCPRGMWGHPTGCPNCNDVSKGPDPCDNKTSGBCHKENHYRPPGSP 1920
DB 1861 YLGPHYCETRIDQPCPRGMWGHPTGCPNCNDVSKGPDPCDNKTSGBCHKENHYRPPGSP 1920
QY 1921 CLCOCYPTFGSLSRVCDPEDGQPCPKRGVYIGROCDRCNPNPABVTTNGCEVAYDSCPAI 1980

```

```

DB 1921 CLCOCYPTFGSLSRVCDPEDGQPCPKRGVYIGROCDRCNPNPABVTTNGCEVAYDSCPAI 1980
QY 1981 EAGIWWPRTFRGLPAAPCPYKSGFSTAVRHCDEHRRGLPBNLFNCTSIPTSELKPAERL 2040
DB 1981 EAGIWWPRTFRGLPAAPCPYKSGFSTAVRHCDEHRRGLPBNLFNCTSIPTSELKPAERL 2040
QY 2041 QNBSGLSGRSQOLALLRNATOTAGYPSGDVAVVQOLATRLAHSTORFGLSATQ 2100
DB 2041 QNBSGLSGRSQOLALLRNATOTAGYPSGDVAVVQOLATRLAHSTORFGLSATQ 2100
QY 2101 DVHFTENILRVGSALLDTANKRMWELIQTGTGTMALLQHEAVASALAKONRHYYLSPF 2160
DB 2101 DVHFTENILRVGSALLDTANKRMWELIQTGTGTMALLQHEAVASALAKONRHYYLSPF 2160
QY 2161 TIVTNTIVISVVRLDKGNFAKAPRYALNGBOPDLETTYILPESVRETTPVVRPAG 2220
DB 2161 TIVTNTIVISVVRLDKGNFAKAPRYALNGBOPDLETTYILPESVRETTPVVRPAG 2220
QY 2221 PGEAOPRELARORRHELSQGEAVASVIYRTLAGLLPHNYDPRKSLRVKAPINT 2280
DB 2221 PGEAOPRELARORRHELSQGEAVASVIYRTLAGLLPHNYDPRKSLRVKAPINT 2280
QY 2281 PVSISVHDEBELPRALDKPVTVQFLLJTEBRTKPICVFMNHSILVSGTGMSARGCE 2340
DB 2281 PVSISVHDEBELPRALDKPVTVQFLLJTEBRTKPICVFMNHSILVSGTGMSARGCE 2340
QY 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENEILLPKLTLYVALGVTLLAALLTFEFTL 2400
DB 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENEILLPKLTLYVALGVTLLAALLTFEFTL 2400
QY 2401 LRILRSNCHIRRLTALGLAQLVFLGIGNOADLPACTYIATILHFLYCTFSWALLE 2460
DB 2401 LRILRSNCHIRRLTALGLAQLVFLGIGNOADLPACTYIATILHFLYCTFSWALLE 2460
QY 2461 ALHLRYALTEYRDVNTGEMRFYMLGWCVPAPITGLAVGLDPEGYNDPCMLSTYDTLI 2520
DB 2461 ALHLRYALTEYRDVNTGEMRFYMLGWCVPAPITGLAVGLDPEGYNDPCMLSTYDTLI 2520
QY 2521 WSPFAPVAFVMSVFLYTLIABASCAAROGFPRKGVSGLOSPFALLLSATWLLAL 2580
DB 2521 WSPFAPVAFVMSVFLYTLIABASCAAROGFPRKGVSGLOSPFALLLSATWLLAL 2580
QY 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLSYVLSKVRKALKLACSRKSPDPALTTKSTL 2640
DB 2581 LSVNSDTLLFHYLPATCNCIOGPFIFLSYVLSKVRKALKLACSRKSPDPALTTKSTL 2640
QY 2641 TTSYNCSPFYADGRLYOPYGDSAGSLHSTSRGKSQPSYIPPLREBSALNPGQPPGLG 2700
DB 2641 TTSYNCSPFYADGRLYOPYGDSAGSLHSTSRGKSQPSYIPPLREBSALNPGQPPGLG 2700
QY 2701 DPGSLFLEGDOQOHPDPTDSDSLSDHDSGYSASTSSSBBEBSBBEBAAPRGOG 2760
DB 2701 DPGSLFLEGDOQOHPDPTDSDSLSDHDSGYSASTSSSBBEBSBBEBAAPRGOG 2760
QY 2761 WDSLILGPABRLPLHSTPKGPGPGKAPMPGADGCTTAKSSGNGAPERLRENGDALSR 2820
DB 2761 WDSLILGPABRLPLHSTPKGPGPGKAPMPGADGCTTAKSSGNGAPERLRENGDALSR 2820
QY 2821 EGSIGPLPGSSAOPHYKGLKKKCLPTISBKSLRLPLBOCTGSSRGSSASBGRGCPP 2880
DB 2821 EGSIGPLPGSSAOPHYKGLKKKCLPTISBKSLRLPLBOCTGSSRGSSASBGRGCPP 2880
QY 2881 RPPPROSLQEOQLNGVMPFAMSIXKATVDEBSSGSEPLPFNFILH 2923
DB 2881 RPPPROSLQEOQLNGVMPFAMSIXKATVDEBSSGSEPLPFNFILH 2923

```

```

RESULT 5
US-10-292-798-932
; Sequence 932, Application us/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:

```



APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: AUBRAUDAN, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
PRIOR FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 932  
LENGTH: 2923  
TYPE: PR  
ORGANISM: Homo sapiens  
US-10-292-798-932

Query Match 100.0%; Score 15545; DB 12; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRSPTATGVP...  
1 MRSPTATGVP...  
61 SASNLWLYSR...  
61 SASNLWLYSR...  
121 GHLSPOGKLT...  
121 GHLSPOGKLT...  
121 GHLSPOGKLT...  
181 POPPSQATV...  
181 POPPSQATV...  
241 VTTAAEELDE...  
241 VTTAAEELDE...  
301 NLEVGVEVL...  
301 NLEVGVEVL...  
361 ESTQVTV...  
361 ESTQVTV...  
421 VTSDDKSG...  
421 VTSDDKSG...  
481 RPLSNVSG...  
481 RPLSNVSG...  
541 EYRLAGV...  
541 EYRLAGV...  
601 LDVNDN...  
601 LDVNDN...  
661 QSGGGV...  
661 QSGGGV...  
721 VNEDRPA...  
721 VNEDRPA...

721 VNEDRPA...  
781 YTLATTA...  
781 YTLATTA...  
841 DEDSG...  
841 DEDSG...  
901 AATPME...  
901 AATPME...  
961 VEGNI...  
961 VEGNI...  
1021 PVLGN...  
1021 PVLGN...  
1081 EKL...  
1081 EKL...  
1141 FLSP...  
1141 FLSP...  
1201 PSED...  
1201 PSED...  
1261 VLF...  
1261 VLF...  
1321 EHC...  
1321 EHC...  
1381 RGL...  
1381 RGL...  
1441 PVP...  
1441 PVP...  
1501 NYS...  
1501 NYS...  
1561 FIAN...  
1561 FIAN...  
1621 GSS...  
1621 GSS...  
1681 GLA...  
1681 GLA...  
1741 NIT...  
1741 NIT...  
1801 PCP...  
1801 PCP...

1801 PCPANSYCSNDMDVSQCDPGYGDNCNVCNLANPCEHQSVCTKRPKSAFHGTCPCPN 1860  
1861 YLGPYCETRIDPCRGWNGHPTGPCNCDVSKGFPDCKNTSGEBCKENHYRPPGSP 1920  
1861 YLGPYCETRIDPCRGWNGHPTGPCNCDVSKGFPDCKNTSGEBCKENHYRPPGSP 1920  
1921 CILCCYCPFGSLSRVDPEDGQCPKPGYIGROCDRCNPPAEVTTNGCEVNYDSCPAI 1980  
1921 CILCCYCPFGSLSRVDPEDGQCPKPGYIGROCDRCNPPAEVTTNGCEVNYDSCPAI 1980  
1921 CILCCYCPFGSLSRVDPEDGQCPKPGYIGROCDRCNPPAEVTTNGCEVNYDSCPAI 1980  
1981 EAGIWMPTRFGLPAAAPCPKSGFGTAVRHCEHRCNMLPNCISITFSEILKGAERL 2040  
1981 EAGIWMPTRFGLPAAAPCPKSGFGTAVRHCEHRCNMLPNCISITFSEILKGAERL 2040  
1981 EAGIWMPTRFGLPAAAPCPKSGFGTAVRHCEHRCNMLPNCISITFSEILKGAERL 2040  
2041 QNBSGLDGRSQOALALLRNATQHTAGYFGSDVKAVALRLLAHSTORFGLSATQ 2100  
2041 QNBSGLDGRSQOALALLRNATQHTAGYFGSDVKAVALRLLAHSTORFGLSATQ 2100  
2101 DVHFTENILRVGSALLDTANKHMLIQOTEGGTAMLQHTAYASALAQNMRHTYLSPP 2160  
2101 DVHFTENILRVGSALLDTANKHMLIQOTEGGTAMLQHTAYASALAQNMRHTYLSPP 2160  
2161 TIVTENVIVSVRLDKGNFAGAKLPRYEALRGOPDLFTTVILPESVRETPPVVRPAG 2220  
2161 TIVTENVIVSVRLDKGNFAGAKLPRYEALRGOPDLFTTVILPESVRETPPVVRPAG 2220  
2221 PGEAOEPEELARORRHPELSOGBAVASIYIYRTLAGLLPHNYDDPKSLRYPKPIINT 2280  
2221 PGEAOEPEELARORRHPELSOGBAVASIYIYRTLAGLLPHNYDDPKSLRYPKPIINT 2280  
2281 PIVSISVHDEELPRALDKPTVQFRLLETERKTPICVFNHNSILVSGTGMSARCE 2340  
2281 PIVSISVHDEELPRALDKPTVQFRLLETERKTPICVFNHNSILVSGTGMSARCE 2340  
2341 VVFRNESHVSCCNHMTSPAVLMDVSRNGBELPKLTLTYVALGVTLLAALLTFEPLTL 2400  
2341 VVFRNESHVSCCNHMTSPAVLMDVSRNGBELPKLTLTYVALGVTLLAALLTFEPLTL 2400  
2401 LRIILSNHGIRRNITAAALGALVFLGIGNADLPACTVAILLHFLYCTFSMALLE 2460  
2401 LRIILSNHGIRRNITAAALGALVFLGIGNADLPACTVAILLHFLYCTFSMALLE 2460  
2461 ALHLRALTEVRDVMGPRFYVMGKGPATITGLAVGLDPBGVGNDFCWLSTYDTLI 2520  
2461 ALHLRALTEVRDVMGPRFYVMGKGPATITGLAVGLDPBGVGNDFCWLSTYDTLI 2520  
2521 MSFAPVAPAVMSVFLYTLARASCAAROGFEKKGPVSGLOPFAVILLISATWLLAL 2580  
2521 MSFAPVAPAVMSVFLYTLARASCAAROGFEKKGPVSGLOPFAVILLISATWLLAL 2580  
2581 LSVNSDTLLFHYLFATCNCIOGPFIFLSYVVLKSVRRALKLACSRKSPDPALTTKSTL 2640  
2581 LSVNSDTLLFHYLFATCNCIOGPFIFLSYVVLKSVRRALKLACSRKSPDPALTTKSTL 2640  
2641 TSSVNCBPVADGRLYOPYGSAGSLHSTRSGKQPSYIIPFLABEBSALNPGQPPGLG 2700  
2641 TSSVNCBPVADGRLYOPYGSAGSLHSTRSGKQPSYIIPFLABEBSALNPGQPPGLG 2700  
2701 DPGSLFLGQDQOHPDTPDSDLGLBDDQSGSYASTSSDSBEEBEEBEEBAAAPGBOG 2760  
2701 DPGSLFLGQDQOHPDTPDSDLGLBDDQSGSYASTSSDSBEEBEEBEEBAAAPGBOG 2760  
2761 WDSLILGPAERLPLHSTPDGPGPGKAPWPGDFGTAKESGNGAPREERLRENDALSR 2820  
2761 WDSLILGPAERLPLHSTPDGPGPGKAPWPGDFGTAKESGNGAPREERLRENDALSR 2820  
2821 EGSIGLPLGSSAOPHKGILKXKCLPTISEKSLALPLBQCTGSRGSSASBGSNGSP 2880  
2821 EGSIGLPLGSSAOPHKGILKXKCLPTISEKSLALPLBQCTGSRGSSASBGSNGSP 2880  
2881 RPPRQSLQOQLNGVPIAMSIKAGTVDEDSGSEFLFNFLLH 2923  
2881 RPPRQSLQOQLNGVPIAMSIKAGTVDEDSGSEFLFNFLLH 2923  
2881 RPPRQSLQOQLNGVPIAMSIKAGTVDEDSGSEFLFNFLLH 2923

RESULT 6  
US-10-225-567A-524  
Sequence 524, Application US/10225567A  
Publication No. US20030113798A1  
GENERAL INFORMATION:  
APPLICANT: Lifespan Biosciences  
APPLICANT: Brown, Joseph P.  
APPLICANT: Burner, Glenna C.  
APPLICANT: Roush, Christine L.  
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 1920-4-4  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/257,144  
PRIOR FILING DATE: 2000-12-19  
NUMBER OF SEQ ID NOS: 2292  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 524  
LENGTH: 2923  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-225-567A-524  
Query Match 100.0%; Score 15545; DB 15; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 MRSPTATGVPPLTPPPPLLLLLLLPPPLLDGQVPCPSLGRSGSSGACAPMGWLCPS 60  
1 MRSPTATGVPPLTPPPPLLLLLLLPPPLLDGQVPCPSLGRSGSSGACAPMGWLCPS 60  
61 SASNMLTYSRCRAGTELGHLPBHDGLVWCPBSBAHPLPAPGCMSCRLIGIG 120  
61 SASNMLTYSRCRAGTELGHLPBHDGLVWCPBSBAHPLPAPGCMSCRLIGIG 120  
121 GHLSPQGLTLPEBHPCLKAPRLRCOSCKLAQAAGLRAERSPEESLGRKRVNTPAQ 180  
121 GHLSPQGLTLPEBHPCLKAPRLRCOSCKLAQAAGLRAERSPEESLGRKRVNTPAQ 180  
181 FQPSYQATVPENOPAGTPVASLALIDPDBGAGRLBYMDALDSSNOFFSLDPVTGA 240  
181 FQPSYQATVPENOPAGTPVASLALIDPDBGAGRLBYMDALDSSNOFFSLDPVTGA 240  
241 VTTAEELDRETKSTHVRVTAADHGMRRSALATLTLYDNTDHPVFBQOYKESLRE 300  
241 VTTAEELDRETKSTHVRVTAADHGMRRSALATLTLYDNTDHPVFBQOYKESLRE 300  
301 NLEVGVEVLTVRATDGDAPPANILYRLBGSQSPSEVEIDPRSGVIRTRGVDBEEV 360  
301 NLEVGVEVLTVRATDGDAPPANILYRLBGSQSPSEVEIDPRSGVIRTRGVDBEEV 360  
361 ESYQUTVEASDQDHPGPRSTTAAVLSVBDNDNAPQFSKRYVVOVREDVTPGAVLR 420  
361 ESYQUTVEASDQDHPGPRSTTAAVLSVBDNDNAPQFSKRYVVOVREDVTPGAVLR 420  
421 VTASDRKGSNAVHYHYSIMGNARGQFYLDAGCALVVSPLDVEETKXTLRYRADGG 480  
421 VTASDRKGSNAVHYHYSIMGNARGQFYLDAGCALVVSPLDVEETKXTLRYRADGG 480  
481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPGLGYLVLAHQALDADAGNARL 540  
481 RPLLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPGLGYLVLAHQALDADAGNARL 540  
541 EYRLAGVGHDPPTINNCTGWSIYAAELDBEEVDYFSGYBAADHKGFPALTASASVTV 600  
541 EYRLAGVGHDPPTINNCTGWSIYAAELDBEEVDYFSGYBAADHKGFPALTASASVTV 600  
601 LDVVDNNPFTQPEYVTRLNEDAAVGVTVTVSAVNDASVITYQTSQNTRRFSITS 660  
601 LDVVDNNPFTQPEYVTRLNEDAAVGVTVTVSAVNDASVITYQTSQNTRRFSITS 660

```

QY 661 QSGGGVSLALPDYKLERQYVLAATASDGTRODTQIYVNTVDANTHAPVFOSSHYYTN 720
Db 661 QSGGGVSLALPDYKLERQYVLAATASDGTRODTQIYVNTVDANTHAPVFOSSHYYTN 720
QY 721 VNEBRPAGTTVVLISATDEDTGENARITYEMEDSIFQFRIDADTGA VTTQALDYEDQVS 780
Db 721 VNEBRPAGTTVVLISATDEDTGENARITYEMEDSIFQFRIDADTGA VTTQALDYEDQVS 780
QY 781 YTLAIFARDNGIPQKSDTTYLBILVNDVNDNAPOFLRDSYQGSVEDVPEFTSVLOISAT 840
Db 781 YTLAIFARDNGIPQKSDTTYLBILVNDVNDNAPOFLRDSYQGSVEDVPEFTSVLOISAT 840
QY 841 DRDSGANGRYFTYFQGGDDGDPPIVESYSGIYRTLRDLRENVAYQVLAAYVNDGMP 900
Db 841 DRDSGANGRYFTYFQGGDDGDPPIVESYSGIYRTLRDLRENVAYQVLAAYVNDGMP 900
QY 901 ARTPEMTVTVLVNDNPPVFEODEPVEEENSPIGLAVARTADPDEGTAAQIMYQI 960
Db 901 ARTPEMTVTVLVNDNPPVFEODEPVEEENSPIGLAVARTADPDEGTAAQIMYQI 960
QY 961 VEGNIDPEVQLOJIFSGELTALVDLYEDREBYVVIQATSAPLVSRATVHRLDENDP 1020
Db 961 VEGNIDPEVQLOJIFSGELTALVDLYEDREBYVVIQATSAPLVSRATVHRLDENDP 1020
QY 1021 PVLGNPEILLNNVYTRSSSPFGGATGRVPAHPDISLSITSPERGENELSTVLNASTG 1080
Db 1021 PVLGNPEILLNNVYTRSSSPFGGATGRVPAHPDISLSITSPERGENELSTVLNASTG 1080
QY 1081 ELKLSALDNNRPLEAIMSVLSDGVSHVTAOCALRVITITDEMLTHSITRLJEDMSPER 1140
Db 1081 ELKLSALDNNRPLEAIMSVLSDGVSHVTAOCALRVITITDEMLTHSITRLJEDMSPER 1140
QY 1141 FLSPLGLPIQAVAAATLAPDPDHVVFNVRQDTDAPEGHILNVSLSVGOPGPGGPPPL 1200
Db 1141 FLSPLGLPIQAVAAATLAPDPDHVVFNVRQDTDAPEGHILNVSLSVGOPGPGGPPPL 1200
QY 1201 PSSDLOERLYLNSLLTASAOVLPDPDNI CLRPEENMRCVSVLRPFSSAPFLASSS 1260
Db 1201 PSSDLOERLYLNSLLTASAOVLPDPDNI CLRPEENMRCVSVLRPFSSAPFLASSS 1260
QY 1261 VLFRPIHPVGLRCRCPGFTGDCYETEVDLCTSRPCPGHRCRSEGGYTCCARDGYTG 1320
Db 1261 VLFRPIHPVGLRCRCPGFTGDCYETEVDLCTSRPCPGHRCRSEGGYTCCARDGYTG 1320
QY 1321 EHCEVSARSGRCTPGVCNKGCTCVNLI VSGPKDCPSGDFEKPYCOVTRSPFAHSPTIF 1380
Db 1321 EHCEVSARSGRCTPGVCNKGCTCVNLI VSGPKDCPSGDFEKPYCOVTRSPFAHSPTIF 1380
QY 1381 RGLRORHFTLALSPATKERDGLLYNGRFNEKHPALAEVIOEVOULTPSAGESTTVS 1440
Db 1381 RGLRORHFTLALSPATKERDGLLYNGRFNEKHPALAEVIOEVOULTPSAGESTTVS 1440
QY 1441 PVPVGVSDQOMHTVOLKYTNKPLAQGTGLPOGSEBQKVA VVVVDGCDTVALRFGSVLG 1500
Db 1441 PVPVGVSDQOMHTVOLKYTNKPLAQGTGLPOGSEBQKVA VVVVDGCDTVALRFGSVLG 1500
QY 1501 NYSCAAGTQGGSKSIDLTGPIILGVPDLPSFPYRMRQFTVGCNKNLOVDSRHIDMAD 1560
Db 1501 NYSCAAGTQGGSKSIDLTGPIILGVPDLPSFPYRMRQFTVGCNKNLOVDSRHIDMAD 1560
QY 1561 FIANNNGVPGCPAKKVCNDNTCHNGGTCYNOMDAFSCBCLPFGGKSCAQEAMNPOHFL 1620
Db 1561 FIANNNGVPGCPAKKVCNDNTCHNGGTCYNOMDAFSCBCLPFGGKSCAQEAMNPOHFL 1620
QY 1621 GSSIVAHGSLPISQPMYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVMLEVEGT 1680
Db 1621 GSSIVAHGSLPISQPMYLSLMFRTQADGVLLQAITRGRSTITTLQLRBGHVMLEVEGT 1680
QY 1681 GLQASSRLRPERGRANDMDHHAOLALGASGPGHAIIISFYGOORABAGNTGPRLLHGHLS 1740
Db 1681 GLQASSRLRPERGRANDMDHHAOLALGASGPGHAIIISFYGOORABAGNTGPRLLHGHLS 1740
QY 1741 NITVGGIPGAGVARGRGLQGVRSVDTPEGVNSLDPSHGESINVEQCSLPDPDCSN 1800
Db 1741 NITVGGIPGAGVARGRGLQGVRSVDTPEGVNSLDPSHGESINVEQCSLPDPDCSN 1800

```

---

```

Db 1741 NITVGGIPGAGVARGRGLQGVRSVDTPEGVNSLDPSHGESINVEQCSLPDPDCSN 1800
QY 1801 PCPANYSNMPDVSQSCDPGYDGNCTNVCNLPCEHOSYCTRKPAFPHGYCECPN 1860
Db 1801 PCPANYSNMPDVSQSCDPGYDGNCTNVCNLPCEHOSYCTRKPAFPHGYCECPN 1860
QY 1861 YLGPYCESTRIDQPCBRGMMGHPCTGPNCDVSKGFPDPCNKTSGBCHCKENHYPGSPPT 1920
Db 1861 YLGPYCESTRIDQPCBRGMMGHPCTGPNCDVSKGFPDPCNKTSGBCHCKENHYPGSPPT 1920
QY 1921 CLICCCYTRGSLSRVCDPDDQCPCKPGVIGRQCRCNPNPAEYVTCNCEVNVYDSCPAI 1980
Db 1921 CLICCCYTRGSLSRVCDPDDQCPCKPGVIGRQCRCNPNPAEYVTCNCEVNVYDSCPAI 1980
QY 1981 EAGIWMRPRFGLPAAACPKGSPGTAVRHCDEHGMPLPMLFNCTSIYFSELKFAERL 2040
Db 1981 EAGIWMRPRFGLPAAACPKGSPGTAVRHCDEHGMPLPMLFNCTSIYFSELKFAERL 2040
QY 2041 QNRESGLDSGSOQLALLRNATQHTAGYFGSDVAVAYOLATRLAHSTORGFLSATQ 2100
Db 2041 QNRESGLDSGSOQLALLRNATQHTAGYFGSDVAVAYOLATRLAHSTORGFLSATQ 2100
QY 2101 DVHFTENILRVGSALLDPANKRMHMLIQOTSGTAMLLQHYEAVASALANMRHTYLSPF 2160
Db 2101 DVHFTENILRVGSALLDPANKRMHMLIQOTSGTAMLLQHYEAVASALANMRHTYLSPF 2160
QY 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALNGEQPDELTYYILPESYFRETPPVVRPAG 2220
Db 2161 TIVTENVIVSVRLDKGNFAGAKLPRYBALNGEQPDELTYYILPESYFRETPPVVRPAG 2220
QY 2221 PGBAOPBEBLARORRHPELSQGBAVASVIYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
Db 2221 PGBAOPBEBLARORRHPELSQGBAVASVIYRTLAGLLPHNYDPDKSLRVPKPIINT 2280
QY 2281 PVVSI SVHDEBELRALDKPVTVOFRLLETEREETKPICVFNHNSILVSGTGMSARGCE 2340
Db 2281 PVVSI SVHDEBELRALDKPVTVOFRLLETEREETKPICVFNHNSILVSGTGMSARGCE 2340
QY 2341 VVFRNESHVSQCNHMTSPA VLMVDSRRNGEILPLKTLTYVALGVTLAALLTFEFLTL 2400
Db 2341 VVFRNESHVSQCNHMTSPA VLMVDSRRNGEILPLKTLTYVALGVTLAALLTFEFLTL 2400
QY 2401 IALISNQHGRNRTALAGLQVFLGINOADI PRACVIAIILHPLYLCTBEMALLE 2460
Db 2401 IALISNQHGRNRTALAGLQVFLGINOADI PRACVIAIILHPLYLCTBEMALLE 2460
QY 2461 ALHLRYALTEVADVNTGPMRFYMLGMGPAPITGAVGLDBEGYGNPFCMLSIYDTLI 2520
Db 2461 ALHLRYALTEVADVNTGPMRFYMLGMGPAPITGAVGLDBEGYGNPFCMLSIYDTLI 2520
QY 2521 WSPFAGVAPAVMSVFLYTLAARASCAAORQGEKKGPVSGLOPSPAVILLISATWLLAL 2580
Db 2521 WSPFAGVAPAVMSVFLYTLAARASCAAORQGEKKGPVSGLOPSPAVILLISATWLLAL 2580
QY 2581 LSVNSDTLLFHTLFATKNCIOQPFITLSYVYLSKEVRLAKLACSRKSPDRLATTKSTL 2640
Db 2581 LSVNSDTLLFHTLFATKNCIOQPFITLSYVYLSKEVRLAKLACSRKSPDRLATTKSTL 2640
QY 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKQPSYIPTLREESALNPGQPPGIG 2700
Db 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTRSGKQPSYIPTLREESALNPGQPPGIG 2700
QY 2701 DPGSLFLBGOQOCHDPDTSDBLSLEDDQSGSVASTHSDBEEBEEBEEBAAFPBGEG 2760
Db 2701 DPGSLFLBGOQOCHDPDTSDBLSLEDDQSGSVASTHSDBEEBEEBEEBAAFPBGEG 2760
QY 2761 WDSILGPGAEPLPLHSTPDGGRGPGKAPWPGDPFGTAKESGNGAPREERLENGDALSR 2820
Db 2761 WDSILGPGAEPLPLHSTPDGGRGPGKAPWPGDPFGTAKESGNGAPREERLENGDALSR 2820
QY 2821 EGSIGPLPGSSAQPHKGIILKKKCLPTISKSLRLPLEOQTGSSRGSASGSRGCPPP 2880
Db 2821 EGSIGPLPGSSAQPHKGIILKKKCLPTISKSLRLPLEOQTGSSRGSASGSRGCPPP 2880

```

```
Db 2821 EGSGLBPSSAQPHKGIKXKCLPTISKSLRLPLEQCTGSSRGSASBSRGGPP 2880
Qy 2881 RPPRQSLQEOJNGVPIAMSIKAGTVDESSGSEFLFNFPLH 2923
Db 2881 RPPRQSLQEOJNGVPIAMSIKAGTVDESSGSEFLFNFPLH 2923

RESULT 7
US-09-788-711A-2
; Sequence 2, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FaSTSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2956
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-788-711A-2

Query Match 99.8%; Score 15518.5; DB 9; Length 2956;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 33; Gaps 1;

Qy 1 MRSPTGVPPLPPPPPLLLLLLLLLPPPLGQVPCSLSGRGGSSGACAPMGLCS 60
Db 1 MRSPTGVPPLPPPPPLLLLLLLLLPPPLGQVPCSLSGRGGSSGACAPMGLCS 60
Qy 61 SASNMLYTSRCRDAGTELTLGHLVPHDGLRVWCPESEAHITLPPAPGSCPMSCLLGIG 120
Db 61 SASNMLYTSRCRDAGTELTLGHLVPHDGLRVWCPESEAHITLPPAPGSCPMSCLLGIG 120
Qy 121 GHLSPQGLTLPEBEHPCLKAPRLRCQSCQLAQPGLRAGENSPEBSLGGRRKRNVTAPQ 180
Db 121 GHLSPQGLTLPEBEHPCLKAPRLRCQSCQLAQPGLRAGENSPEBSLGGRRKRNVTAPQ 180
Qy 121 GHLSPQGLTLPEBEHPCLKAPRLRCQSCQLAQPGLRAGENSPEBSLGGRRKRNVTAPQ 180
Db 121 GHLSPQGLTLPEBEHPCLKAPRLRCQSCQLAQPGLRAGENSPEBSLGGRRKRNVTAPQ 180
Qy 181 POPBPYQATVPENOGAGTPVASTRAIDPDEGAGLETMTMLPFSRNSQFSLDPVTGA 240
Db 181 POPBPYQATVPENOGAGTPVASTRAIDPDEGAGLETMTMLPFSRNSQFSLDPVTGA 240
Qy 181 POPBPYQATVPENOGAGTPVASTRAIDPDEGAGLETMTMLPFSRNSQFSLDPVTGA 240
Db 181 POPBPYQATVPENOGAGTPVASTRAIDPDEGAGLETMTMLPFSRNSQFSLDPVTGA 240
Qy 241 VTTAEELRETKSTVFRVTAQDHGMPRSALATLTILVTDTNDHPVFEQOEYKESLRE 300
Db 241 VTTAEELRETKSTVFRVTAQDHGMPRSALATLTILVTDTNDHPVFEQOEYKESLRE 300
Qy 241 VTTAEELRETKSTVFRVTAQDHGMPRSALATLTILVTDTNDHPVFEQOEYKESLRE 300
Db 241 VTTAEELRETKSTVFRVTAQDHGMPRSALATLTILVTDTNDHPVFEQOEYKESLRE 300
Qy 301 NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFEIDPRSGVIRTRGPVDRBEV 360
Db 301 NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFEIDPRSGVIRTRGPVDRBEV 360
Qy 301 NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFEIDPRSGVIRTRGPVDRBEV 360
Db 301 NLEVGVEVLTVAATDGDAPPNANILYRLLEGSGSPSEVFEIDPRSGVIRTRGPVDRBEV 360
Qy 361 ESYQULTVASODGRPGPRSTTAAYFLSYEDNDNAPQSEKRYVQVREDVTPCAPVLK 420
Db 361 ESYQULTVASODGRPGPRSTTAAYFLSYEDNDNAPQSEKRYVQVREDVTPCAPVLK 420
Qy 361 ESYQULTVASODGRPGPRSTTAAYFLSYEDNDNAPQSEKRYVQVREDVTPCAPVLK 420
Db 361 ESYQULTVASODGRPGPRSTTAAYFLSYEDNDNAPQSEKRYVQVREDVTPCAPVLK 420
Qy 421 VTASDRDKSNANVHYSIMSGNARCOFYLDAGTALDVYSPADYETTKSEYTLKRVADGG 480
Db 421 VTASDRDKSNANVHYSIMSGNARCOFYLDAGTALDVYSPADYETTKSEYTLKRVADGG 480
Qy 421 VTASDRDKSNANVHYSIMSGNARCOFYLDAGTALDVYSPADYETTKSEYTLKRVADGG 480
Db 421 VTASDRDKSNANVHYSIMSGNARCOFYLDAGTALDVYSPADYETTKSEYTLKRVADGG 480
Qy 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALIDADAGDNRL 540
Db 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALIDADAGDNRL 540
Qy 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALIDADAGDNRL 540
Db 481 RPPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESVPLGYLVLAHQALIDADAGDNRL 540
Qy 541 EYRLAGVGHDPPTLNNNGMTSVAAEILREVDVYSRGVTRBDGTPALTSASVSTV 600
Db 541 EYRLAGVGHDPPTLNNNGMTSVAAEILREVDVYSRGVTRBDGTPALTSASVSTV 600
Qy 541 EYRLAGVGHDPPTLNNNGMTSVAAEILREVDVYSRGVTRBDGTPALTSASVSTV 600
Db 541 EYRLAGVGHDPPTLNNNGMTSVAAEILREVDVYSRGVTRBDGTPALTSASVSTV 600
Qy 601 LDVNNNPTFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSYITVQITSGNTRNFSITS 660
Db 601 LDVNNNPTFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSYITVQITSGNTRNFSITS 660

Db 601 LDVNNNPTFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSYITVQITSGNTRNFSITS 660
Qy 661 OSGGGLVSLALPLDYKLERQVYLAVTASDGTRODTAQIVANNVTANTHRPVQSSHYTVN 720
Db 661 OSGGGLVSLALPLDYKLERQVYLAVTASDGTRODTAQIVANNVTANTHRPVQSSHYTVN 720
Qy 721 VNEBRPAGITTVVLSATDEDTGENARITYFMEDSIPOFRIDAGTAVTQALDYEDQVS 780
Db 721 VNEBRPAGITTVVLSATDEDTGENARITYFMEDSIPOFRIDAGTAVTQALDYEDQVS 780
Qy 721 VNEBRPAGITTVVLSATDEDTGENARITYFMEDSIPOFRIDAGTAVTQALDYEDQVS 780
Db 721 VNEBRPAGITTVVLSATDEDTGENARITYFMEDSIPOFRIDAGTAVTQALDYEDQVS 780
Qy 781 YTLATTAANDNGIPQKSDTYLIELVNDVNDNAPQFLRDSYQGSYVEDVPPFTSVQISAT 840
Db 781 YTLATTAANDNGIPQKSDTYLIELVNDVNDNAPQFLRDSYQGSYVEDVPPFTSVQISAT 840
Qy 781 YTLATTAANDNGIPQKSDTYLIELVNDVNDNAPQFLRDSYQGSYVEDVPPFTSVQISAT 840
Db 781 YTLATTAANDNGIPQKSDTYLIELVNDVNDNAPQFLRDSYQGSYVEDVPPFTSVQISAT 840
Qy 841 DRDSGLANGRVFTYFGGDDGDGDFIVESTSGIVRLRLRDEBNVAQVYLAADVKGMP 900
Db 841 DRDSGLANGRVFTYFGGDDGDGDFIVESTSGIVRLRLRDEBNVAQVYLAADVKGMP 900
Qy 841 DRDSGLANGRVFTYFGGDDGDGDFIVESTSGIVRLRLRDEBNVAQVYLAADVKGMP 900
Db 841 DRDSGLANGRVFTYFGGDDGDGDFIVESTSGIVRLRLRDEBNVAQVYLAADVKGMP 900
Qy 901 ARTPMEVTVTVLDVNDNPPVFEQDEPFVFEENSPIGLAVARVTATDDEGTNAQIMYQI 960
Db 901 ARTPMEVTVTVLDVNDNPPVFEQDEPFVFEENSPIGLAVARVTATDDEGTNAQIMYQI 960
Qy 901 ARTPMEVTVTVLDVNDNPPVFEQDEPFVFEENSPIGLAVARVTATDDEGTNAQIMYQI 960
Db 901 ARTPMEVTVTVLDVNDNPPVFEQDEPFVFEENSPIGLAVARVTATDDEGTNAQIMYQI 960
Qy 961 VEGNIPVEYFOLDIFSGLTALVDLDYEDRPEYVLTQATSAPIVSRAVTVHRLDRNNP 1020
Db 961 VEGNIPVEYFOLDIFSGLTALVDLDYEDRPEYVLTQATSAPIVSRAVTVHRLDRNNP 1020
Qy 961 VEGNIPVEYFOLDIFSGLTALVDLDYEDRPEYVLTQATSAPIVSRAVTVHRLDRNNP 1020
Db 961 VEGNIPVEYFOLDIFSGLTALVDLDYEDRPEYVLTQATSAPIVSRAVTVHRLDRNNP 1020
Qy 1021 PVLGNFELFNNNYNNSSSPFGAIGRVPAHDPDISDLYTSFERGNELSLVILNASTG 1080
Db 1021 PVLGNFELFNNNYNNSSSPFGAIGRVPAHDPDISDLYTSFERGNELSLVILNASTG 1080
Qy 1021 PVLGNFELFNNNYNNSSSPFGAIGRVPAHDPDISDLYTSFERGNELSLVILNASTG 1080
Db 1021 PVLGNFELFNNNYNNSSSPFGAIGRVPAHDPDISDLYTSFERGNELSLVILNASTG 1080
Qy 1081 ELKLSRALDNNRPLBAIMSVTVSDGVSVTQAQALRVITITDEMUTHSITTLRLDMSPER 1140
Db 1081 ELKLSRALDNNRPLBAIMSVTVSDGVSVTQAQALRVITITDEMUTHSITTLRLDMSPER 1140
Qy 1081 ELKLSRALDNNRPLBAIMSVTVSDGVSVTQAQALRVITITDEMUTHSITTLRLDMSPER 1140
Db 1081 ELKLSRALDNNRPLBAIMSVTVSDGVSVTQAQALRVITITDEMUTHSITTLRLDMSPER 1140
Qy 1141 FLSPILGFIQAAVATLATPPDHVVNVNVRDTPAPGSHILNVLISVQGPBGSGPFL 1200
Db 1141 FLSPILGFIQAAVATLATPPDHVVNVNVRDTPAPGSHILNVLISVQGPBGSGPFL 1200
Qy 1141 FLSPILGFIQAAVATLATPPDHVVNVNVRDTPAPGSHILNVLISVQGPBGSGPFL 1200
Db 1141 FLSPILGFIQAAVATLATPPDHVVNVNVRDTPAPGSHILNVLISVQGPBGSGPFL 1200
Qy 1201 PSEDLQERLYNRSILTAISAQRYLPDDNLCLEBPCENYRQVSLRFDSSAPFIASS 1260
Db 1201 PSEDLQERLYNRSILTAISAQRYLPDDNLCLEBPCENYRQVSLRFDSSAPFIASS 1260
Qy 1201 PSEDLQERLYNRSILTAISAQRYLPDDNLCLEBPCENYRQVSLRFDSSAPFIASS 1260
Db 1201 PSEDLQERLYNRSILTAISAQRYLPDDNLCLEBPCENYRQVSLRFDSSAPFIASS 1260
Qy 1261 VLFRPIHPVGLRRCRCPGFTGDYCEYEVLDLQYRPGCPHRCRSRGGYTCRLRDGTG 1320
Db 1261 VLFRPIHPVGLRRCRCPGFTGDYCEYEVLDLQYRPGCPHRCRSRGGYTCRLRDGTG 1320
Qy 1261 VLFRPIHPVGLRRCRCPGFTGDYCEYEVLDLQYRPGCPHRCRSRGGYTCRLRDGTG 1320
Db 1261 VLFRPIHPVGLRRCRCPGFTGDYCEYEVLDLQYRPGCPHRCRSRGGYTCRLRDGTG 1320
Qy 1321 EHCVSARSRCCTPGVCNKGTCYNLLVYGFKCCPSPGDFEKPVCQVYTRSFPAHSFTF 1380
Db 1321 EHCVSARSRCCTPGVCNKGTCYNLLVYGFKCCPSPGDFEKPVCQVYTRSFPAHSFTF 1380
Qy 1321 EHCVSARSRCCTPGVCNKGTCYNLLVYGFKCCPSPGDFEKPVCQVYTRSFPAHSFTF 1380
Db 1321 EHCVSARSRCCTPGVCNKGTCYNLLVYGFKCCPSPGDFEKPVCQVYTRSFPAHSFTF 1380
Qy 1381 RGLRQRFHFTLALSPTKBRDGLLYNGRFNEKHDVVALVYLOBOVLTFSAGSSTTVS 1440
Db 1381 RGLRQRFHFTLALSPTKBRDGLLYNGRFNEKHDVVALVYLOBOVLTFSAGSSTTVS 1440
Qy 1381 RGLRQRFHFTLALSPTKBRDGLLYNGRFNEKHDVVALVYLOBOVLTFSAGSSTTVS 1440
Db 1381 RGLRQRFHFTLALSPTKBRDGLLYNGRFNEKHDVVALVYLOBOVLTFSAGSSTTVS 1440
Qy 1441 PFPVPGVSDGQWHTVQLKTYNKPILLGOTGLPQGSSEQKVAVVTYDGDCTYALAFSGSLG 1500
Db 1441 PFPVPGVSDGQWHTVQLKTYNKPILLGOTGLPQGSSEQKVAVVTYDGDCTYALAFSGSLG 1500
Qy 1441 PFPVPGVSDGQWHTVQLKTYNKPILLGOTGLPQGSSEQKVAVVTYDGDCTYALAFSGSLG 1500
Db 1441 PFPVPGVSDGQWHTVQLKTYNKPILLGOTGLPQGSSEQKVAVVTYDGDCTYALAFSGSLG 1500
Qy 1501 NYSQAQOTQGSKKSLDLNGLPILLGGVPLDESEFPARMROFVGCNMLQVDSHIDMAD 1560
Db 1501 NYSQAQOTQGSKKSLDLNGLPILLGGVPLDESEFPARMROFVGCNMLQVDSHIDMAD 1560
Qy 1501 NYSQAQOTQGSKKSLDLNGLPILLGGVPLDESEFPARMROFVGCNMLQVDSHIDMAD 1560
Db 1501 NYSQAQOTQGSKKSLDLNGLPILLGGVPLDESEFPARMROFVGCNMLQVDSHIDMAD 1560
Qy 1561 FIANNGVPCGPAAKNVCDNSTCANCGTCVNMQDAFSCBCEPLGFGKSCAQEMANPQIFL 1620
Db 1561 FIANNGVPCGPAAKNVCDNSTCANCGTCVNMQDAFSCBCEPLGFGKSCAQEMANPQIFL 1620
Qy 1561 FIANNGVPCGPAAKNVCDNSTCANCGTCVNMQDAFSCBCEPLGFGKSCAQEMANPQIFL 1620
Db 1561 FIANNGVPCGPAAKNVCDNSTCANCGTCVNMQDAFSCBCEPLGFGKSCAQEMANPQIFL 1620
Qy 1621 GSSLVAMHGLSLPISQPMYLSIMFRTOADGVLLQALTRGSTTTLQIRBSHVLSEGT 1680
Db 1621 GSSLVAMHGLSLPISQPMYLSIMFRTOADGVLLQALTRGSTTTLQIRBSHVLSEGT 1680
Qy 1621 GSSLVAMHGLSLPISQPMYLSIMFRTOADGVLLQALTRGSTTTLQIRBSHVLSEGT 1680
Db 1621 GSSLVAMHGLSLPISQPMYLSIMFRTOADGVLLQALTRGSTTTLQIRBSHVLSEGT 1680
Qy 1681 GLQASSLRLEPGRANDDMHAQALAGAGGPAHLSFDVGOQARAGNGLPRLHGLHS 1740
Db 1681 GLQASSLRLEPGRANDDMHAQALAGAGGPAHLSFDVGOQARAGNGLPRLHGLHS 1740
Qy 1681 GLQASSLRLEPGRANDDMHAQALAGAGGPAHLSFDVGOQARAGNGLPRLHGLHS 1740
Db 1681 GLQASSLRLEPGRANDDMHAQALAGAGGPAHLSFDVGOQARAGNGLPRLHGLHS 1740
```

QY 1741 NITVGGIIPGAGVARGFRGLQGVAVSDTEEGVNSLDPESHGESINVEQCSLPDPCDSN 1800  
 Db 1741 NITVGGIIPGAGVARGFRGLQGVAVSDTEEGVNSLDPESHGESINVEQCSLPDPCDSN 1800  
 QY 1801 PCPANSYCSNDWMSYSCSCDPGYTGDNCTVCDLNPCEHOSVCTKPSAPHGTYTCCPCPN 1860  
 Db 1801 PCPANSYCSNDWMSYSCSCDPGYTGDNCTVCDLNPCEHOSVCTKPSAPHGTYTCCPCPN 1860  
 QY 1861 YLGPYCTRIIDQPCPRGMWGHPTGPGNCVDYSKGFDDCNKTSGECHCKENHTRPPSPPT 1920  
 Db 1861 YLGPYCTRIIDQPCPRGMWGHPTGPGNCVDYSKGFDDCNKTSGECHCKENHTRPPSPPT 1920  
 QY 1921 CLLCDCTPTGSLASRCDPEDQCPCKEGLVGRQCDRCNDPFAEYTTNGCEVNTYDSCPRAI 1980  
 Db 1921 CLLCDCTPTGSLASRCDPEDQCPCKEGLVGRQCDRCNDPFAEYTTNGCEVNTYDSCPRAI 1980  
 QY 1981 EAGIWMFRTFRGLPAAAPCPKPSGRTAVRHCDERHGMPLPMLFNCTISITSEBLKGFARL 2040  
 Db 1981 EAGIWMFRTFRGLPAAAPCPKPSGRTAVRHCDERHGMPLPMLFNCTISITSEBLKGFARL 2040  
 QY 2041 QRNBSGLDSGRSQDLALLRNATQHTAGYFGSDVKAVYOLATRLAHSTORGFGLSATQ 2100  
 Db 2041 QRNBSGLDSGRSQDLALLRNATQHTAGYFGSDVKAVYOLATRLAHSTORGFGLSATQ 2100  
 QY 2101 DVHFTENLAVGSALDPTANKRHWELIQTEGGTAMLOHYEAVASALAQNMHTYLSPT 2160  
 Db 2101 DVHFTENLAVGSALDPTANKRHWELIQTEGGTAMLOHYEAVASALAQNMHTYLSPT 2160  
 QY 2161 TTYTPNIVISVNRDKNFPAGAKLPRYBALRGEOPDELTTVILPESVFETPPVVRPAG 2220  
 Db 2161 TTYTPNIVISVNRDKNFPAGAKLPRYBALRGEOPDELTTVILPESVFETPPVVRPAG 2220  
 QY 2221 PGBAQEBELARQRRPELSQGEAVASVILYRTLAGLPHANDPDGRSARVPRPIINT 2280  
 Db 2221 PGBAQEBELARQRRPELSQGEAVASVILYRTLAGLPHANDPDGRSARVPRPIINT 2280  
 QY 2281 PVSISVHDEBELPRALDPRVTVOFLLEETERTKICVFMNHSILVSGTGMASARGC 2340  
 Db 2281 PVSISVHDEBELPRALDPRVTVOFLLEETERTKICVFMNHSILVSGTGMASARGC 2340  
 QY 2341 VVFRNESHVSQCCNHNMTSPAVLMDVSRREGEILPLTLYVVALGVLALLLTFFFLTL 2400  
 Db 2341 VVFRNESHVSQCCNHNMTSPAVLMDVSRREGEILPLTLYVVALGVLALLLTFFFLTL 2400  
 QY 2401 LRILRSNQHIRRNLTALGLAOLVFLGLINQADLPACTVIALHLFLYLCFFSWALLE 2460  
 Db 2401 LRILRSNQHIRRNLTALGLAOLVFLGLINQADLPACTVIALHLFLYLCFFSWALLE 2460  
 QY 2461 ALHLRYALTEVRDVTGPMRFYTMLGWVPAFTTGLAVGLDPGKGNDFCWLSTYDTLI 2520  
 Db 2461 ALHLRYALTEVRDVTGPMRFYTMLGWVPAFTTGLAVGLDPGKGNDFCWLSTYDTLI 2520  
 QY 2521 WSPRGPAPAFVMSVFLYTLAARASCAARQGEKGVSGLOPSPFVALLLSATWLLAL 2580  
 Db 2521 WSPRGPAPAFVMSVFLYTLAARASCAARQGEKGVSGLOPSPFVALLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHYLFATCNCTIGPFIPLSYVVLSEKRYAKLACSRKSPDPALTTKSTL 2640  
 Db 2581 LSVNSDTLLFHYLFATCNCTIGPFIPLSYVVLSEKRYAKLACSRKSPDPALTTKSTL 2640  
 QY 2641 TTSSTNCPSPYADGRLVOPYDASAGLSHSTRSGKSPSYPLILREESALNPGQPPGGLG 2700  
 Db 2641 TTSSTNCPSPYADGRLVOPYDASAGLSHSTRSGKSPSYPLILREESALNPGQPPGGLG 2700  
 QY 2701 DPGSLFLEGDOOHDPPTDSDSLLEDDQSGVASTHSSEBEEBEEBBAAPGBOG 2760  
 Db 2701 DPGSLFLEGDOOHDPPTDSDSLLEDDQSGVASTHSSEBEEBEEBBAAPGBOG 2760  
 QY 2761 WDSLIGGARLPPLHSTPKDQGPGRKAPWPGDFTTAKESSGNGAPBEERLRENGDALSR 2820  
 Db 2761 WDSLIGGARLPPLHSTPKDQGPGRKAPWPGDFTTAKESSGNGAPBEERLRENGDALSR 2820

QY 2821 EGSLGPLPSSSAQPHK-----GILKKCLPPT 2847  
 Db 2821 EGSLGPLPSSSAQPHKGEWGTSPCRAPLVSLIPLHMLPHTAPRPRGILKKCLPPT 2880  
 QY 2848 SEKSLRLPLEBCTSSSGSSASBSGRRGPPRRPPROSLQOLNGVPIAMSIAGTY 2907  
 Db 2881 SEKSLRLPLEBCTSSSGSSASBSGRRGPPRRPPROSLQOLNGVPIAMSIAGTY 2940  
 QY 2908 DEDSGSEFLFNFVFLH 2923  
 Db 2941 DEDSGSEFLFNFVFLH 2956  
 RESULT 8  
 US-09-737-149-25  
 ; Sequence 25, Application US/09737149  
 ; Patent No. US20020077466A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Spaderna, Steven K  
 ; APPLICANT: Quinn, Kerry E.  
 ; APPLICANT: Shinkets, Richard A.  
 ; APPLICANT: Muralidhara, Padigaru  
 ; APPLICANT: Spytek, Kimberly A.  
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 15966-620 CIP  
 ; CURRENT APPLICATION NUMBER: US/09/737,149  
 ; PRIOR FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: 60/170,564  
 ; PRIOR FILING DATE: 1999-12-14  
 ; PRIOR APPLICATION NUMBER: 60/173,165  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: 60/173,362  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: 60/173,544  
 ; PRIOR FILING DATE: 1999-12-29  
 ; PRIOR APPLICATION NUMBER: 60/174,404  
 ; PRIOR FILING DATE: 2000-01-04  
 ; PRIOR APPLICATION NUMBER: 60/174,962  
 ; PRIOR FILING DATE: 2000-01-07  
 ; PRIOR APPLICATION NUMBER: 60/223,929  
 ; PRIOR FILING DATE: 2000-08-09  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 25  
 ; LENGTH: 3034  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-737-149-25  
 Query Match 57.7%; Score 8974.5; DB 9; Length 3034;  
 Best Local Similarity 57.3%; Pred. No. 0;  
 Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;  
 QY 1 MSPRATGVLPFPPEPPLLLLLLPPPLGQVQPCSLSGRGS-----S 48  
 Db 124 LRSRAGAELRSP-----AVRSVPLGDAL--CPRAGGGAASLTSVLEAITNPPA 172  
 QY 49 GACAPM-----GWLCC--PSSASMLYTSRCDAGTELTGHLVPHHDGLVWCPSEBA 99  
 Db 173 CCCEPVAGCGCRGRPICLRPGSABELRLVCAIGRAAG-----VWV----- 213  
 QY 100 HPLPAPBPCWSCRLIGTGHLSPOGKLTLPBEPCLKAPRLRCQSCKLQADGLRAG 159  
 Db 214 -----ELVYQATSGTPSPSPV-SPSL-----LNLGQP--RAG 243  
 QY 160 --ERSPEBSLGRKRKNVNTAPQOPPSYQATVPENOPAGTVASLRAIDPDEGEAGLE 217  
 Db 244 VVRKS-----RKGTGSSSTSPQFPUPSTQVSFENRPGTVITELRAHPDEBGDGRLS 296  
 QY 218 YTMALFDRSRNQPSLDPVTGAVTTABELDRBTKSTHFRVTAQDHGMRRSALATLTI 277  
 Db 297 YQMBALFDRSRNGYFLIDATGAVTTABELDRBTKSTHFRVTAQDHGMRRSALATLTI 356

QY 278 LVTDMNDPVRBOEYKESLREMLEVGEVLTTRATCDADPANNILYRLBGGSGSPS 337  
DB 357 TVSDTNDHSPVFEQOSREIREREMLEVEGEVLTIRATDGDAPSNAMRYRLBAGG--- 413  
QY 338 BYFEIDPBRGVI RTBSPVDRBEVSEYQOLTEASDQDRDGPSTTAAPVLASVEDDNDAP 397  
DB 414 -VFEIDANSGVVRTRAVNREBEAEYQOLTEANDQGRNPGLSASATVHI VVEDENDNP 472  
QY 398 QSEKRYVVOVEDVTPGAPVLRTASDRDKGSNAVHVYSIMSGNARQGPYLDQAOTGALD 457  
DB 472 QSEKRYVVOVEDVTPGAPVLRTASDRDKGSNAVHVYSIMSGNARQGPYLDQAOTGALD 532  
QY 458 VSPDLDEYTKYTLRVARQDGRPPLSVNGSLVTVQVLDINDNAPIVYSTPPQATLES 517  
DB 533 VINPDLFEAIRREYTLRIKADGGRPPLINSSGLVSVQVLDVNDNAPIFVSSPFOALVLEN 592  
QY 518 VPLGYLVHVOAIDADAGNARLEAYLAGVH-----DPEPTINNGG 560  
DB 593 VPLGSHVTHIQAVDADAGENARLQYRLVDTASTYVGGSSVDSBNPASADPPPOHNSG 652  
QY 561 WISVAELDRBEVDPYSPGVAEADHGTPALITASASVTVLDVNDNPTFTQPEYTVRLN 620  
DB 653 WITVCAELDRBEVDPYSPGVAEADHGTPALITASASVTVLDVNDNPTFTQPEYTVRLN 712  
QY 621 EDAAVGTSVTVVSAVDRDAHSVITTOITSGNTRNRPSTISQSGGLVSLALPLDYKLERQ 680  
DB 713 EDAAVGTSVTVVSAVDRDAHSVITTOITSGNTRNRPSTISQSGGLVSLALPLDYKLERQ 772  
QY 681 YLAVATASDGTQDPTAOIVVNTDANTHRPVFQSSHYTNNMEDRPACTVVLISATDSD 740  
DB 773 YLAVATASDGTQDPTAOIVVNTDANTHRPVFQSSHYTNNMEDRPACTVVLISATDSD 832  
QY 741 TGENARITYFEMDSIPQFRIADDTGAVTTQAEILDYEDQVSYTLATARDNGIPQKSDTTY 800  
DB 833 TGENARITYFEMDSIPQFRIADDTGAVTTQAEILDYEDQVSYTLATARDNGIPQKSDTTY 892  
QY 801 LEILVNDVNDNAPOPLRDSYQGSVEDVPPPTSVLQISATDSDSGLNGVPTTPOGGDGG 860  
DB 893 LEILVNDVNDNAPOPLRDSYQGSVEDVPPPTSVLQISATDSDSGLNGVPTTPOGGDGG 952  
QY 861 DDDPFIVESTGIVRTLRRLDRRENVAAQYVLRAYAVXKGM-PARTIMEVTVTVLDVNDNP 919  
DB 953 DDDPFIVESTGIVRTLRRLDRRENVAAQYVLRAYAVXKGM-PARTIMEVTVTVLDVNDNP 1012  
QY 920 VEBODEFDVFEVENSEIGLAVAVTATDDEGTNAQIMQIYBGNPEVFOLDIFSGELT 979  
DB 1013 VEBODEFDVFEVENSEIGLAVAVTATDDEGTNAQIMQIYBGNPEVFOLDIFSGELT 1072  
QY 980 ALVYDDYERPREVTVYQATSAPLYSRATVAVRLDRDNPVVLGNFELFNNYVNNSS 1039  
DB 1073 ALVYDDYERPREVTVYQATSAPLYSRATVAVRLDRDNPVVLGNFELFNNYVNNSS 1132  
QY 1040 SPFGAGIGVPAHDDPISLSLYSPERNGELSVLNLASTGELKSRALDNNRPLEAIMS 1099  
DB 1133 SPFGAGIGVPAHDDPISLSLYSPERNGELSVLNLASTGELKSRALDNNRPLEAIMS 1192  
QY 1100 VLVSDGVSHVTAQALRVITITDEMILTHSITLRLBEMSPERFASPLGLHIOAVATLAT 1159  
DB 1193 VLVSDGVSHVTAQALRVITITDEMILTHSITLRLBEMSPERFASPLGLHIOAVATLAT 1252  
QY 1160 PRDHAIVVAVQDTPAPGHHIIVNSLSVQCPQPGGPPPLPSEDIQERLYANRSLTNI 1219  
DB 1253 PRDHAIVVAVQDTPAPGHHIIVNSLSVQCPQPGGPPPLPSEDIQERLYANRSLTNI 1309  
QY 1220 SAORVLPFPDNI CLRPCENYRCSVVLRFPSSAPFIASSVLPFRPHVGLRRCRCPG 1279  
DB 1310 SAORVLPFPDNI CLRPCENYRCSVVLRFPSSAPFIASSVLPFRPHVGLRRCRCPG 1369  
QY 1280 FTGDYCEFEVDLCYSRPCGPHRCRSREGGYTCLCRDGYTGHCEVANSRGCTPGVCN 1339  
DB 1370 FTGDYCEFEVDLCYSRPCGPHRCRSREGGYTCLCRDGYTGHCEVANSRGCTPGVCN 1429  
QY 1340 GGTGVNLVGVKCCDPSGDPEKPYCQVTTTRSPFAHSPIITFRGLRORFHTLALSFAIK 1399

DB 1430 GGTGVNLVGVKCCDPSGEYEHPCYCEVSTRSPFPQSVFTRFGLRORFHTLALSFAIOTD 1489  
QY 1400 RDGILLNCRFBKBDPALAVIOQVOULTSAGSTTTTSPFVPGCVSDQGMHVOLKY 1459  
DB 1490 RUALLVNCRFBKBDPALAVIOQVOULTSAGSTTTTSPFVPGCVSDQGMHVOLKY 1549  
QY 1460 YNKPLIGOTGLPQGSBQKAVVAVVWDGDTGVALRFGSVLGNVSCAAGTOGSGKSIDL 1519  
DB 1550 YNKPLIGOTGLPQGSBQKAVVAVVWDGDTGVALRFGSVLGNVSCAAGTOGSGKSIDL 1609  
QY 1520 TGPLLIGVPLDPSFPYRMRQVYGCNENLOVDSHHIMADFIANGTVPGCPARKAVCD 1579  
DB 1610 TGPLLIGVPLDPSFPYRMRQVYGCNENLOVDSHHIMADFIANGTVPGCPARKAVCD 1669  
QY 1580 SNTCHNGGTCNOMDAPSCBCPLRGKSGCAQEMANPQHPLGSSLVANHGSLPISQWY 1639  
DB 1670 SNTCHNGGTCNOMDAPSCBCPLRGKSGCAQEMANPQHPLGSSLVANHGSLPISQWY 1729  
QY 1640 LSLMFRTOADGVLLQAITRGRSTTTQLRBGRVWLVEGTGLQASLRLBGRANDGDW 1699  
DB 1730 LSLMFRTOADGVLLQAITRGRSTTTQLRBGRVWLVEGTGLQASLRLBGRANDGDW 1789  
QY 1700 HHAQIALASGAPGH-----ALISFDYGOQARAKNLGRLHGLHSNTVGGIPGPAGV 1754  
DB 1790 HHAQIALASGAPGH-----ALISFDYGOQARAKNLGRLHGLHSNTVGGIPGPAGV 1848  
QY 1755 ARGFRGLQGVRSPTPBGVNSLDBSHBSINVQCGSLPPOCSNPPCANYSYCNMDS 1814  
DB 1849 ARGFRGLQGVRSPTPBGVNSLDBSHBSINVQCGSLPPOCSNPPCANYSYCNMDS 1908  
QY 1815 YSCSDPGEYDNCNVCNVDLNPCEHQSVCTRKPSAPHGYCECPNYPGAPCETRIDPC 1874  
DB 1909 YSCSDPGEYDNCNVCNVDLNPCEHQSVCTRKPSAPHGYCECPNYPGAPCETRIDPC 1968  
QY 1875 PRGMWGHPTGCP CUCDVSKGFPDPCNKTSGECHKENHYRPPGSPFTCLLCCTPYGSLR 1934  
DB 1969 PRGMWGHPTGCP CUCDVSKGFPDPCNKTSGECHKENHYRPPGSPFTCLLCCTPYGSLR 2028  
QY 1935 VCDPEDGQCPCKPGVYIGRCQRCNPPFAVYTNNGCVVYDSCPRILEGIMWPRFRPLP 1994  
DB 2029 VCDPEDGQCPCKPGVYIGRCQRCNPPFAVYTNNGCVVYDSCPRILEGIMWPRFRPLP 2088  
QY 1995 AAAPCPKGSFGTAIVAHCDERHGMPLPMLFNCTSTFSEILKFAERLQHNESGLDSGRSQ 2054  
DB 2089 AAAPCPKGSFGTAIVAHCDERHGMPLPMLFNCTSTFSEILKFAERLQHNESGLDSGRSQ 2148  
QY 2055 LALLIRNATQNSTLFGNDVAVTAYQOLARILQHSRQGFPLAATREANFEDVHTGSA 2114  
DB 2149 LALLIRNATQNSTLFGNDVAVTAYQOLARILQHSRQGFPLAATREANFEDVHTGSA 2208  
QY 2115 LLDPTANKHWEILQOTBEGTAMLOHYEAVASALAKONNRHLYLSPFTVTVNIVSVRL 2174  
DB 2209 LLDPTANKHWEILQOTBEGTAMLOHYEAVASALAKONNRHLYLSPFTVTVNIVSVRL 2268  
QY 2175 DKGNFAGKCLPRYALNKGQPPDLETTVILPESVF-----RETPPVVR-----PAGE 2221  
DB 2269 DKGNFAGKCLPRYALNKGQPPDLETTVILPESVF-----RETPPVVR-----PAGE 2328  
QY 2222 GBAQPEELARORRHPBELSGEAVASYITRTLAGLLPHNYDDPKSLRVPKPIINTP 2281  
DB 2329 GBAQPEELARORRHPBELSGEAVASYITRTLAGLLPHNYDDPKSLRVPKPIINTP 2388  
QY 2282 VVSIIVHDEDELPAALDKPVTYVFPRLLETERKPICVFNHNSILVSGTGMARSAGEV 2341  
DB 2389 VVSIIVHDEDELPAALDKPVTYVFPRLLETERKPICVFNHNSILVSGTGMARSAGEV 2448  
QY 2342 VFNESHVSQCNEMTSFAVLMADVSRRENGEILDKLTLYVALAALLTFFFTLTL 2401  
DB 2449 VFNESHVSQCNEMTSFAVLMADVSRRENGEILDKLTLYVALAALLTFFFTLTL 2508  
QY 2402 RILSNQHGIRNNTALALGLAOLVFLGLGINQADLPACTVATILHPIYLCTFSMALIEA 2461



DB 2509 RTLRSLHSIHKNLIALPFSOLI PMVGINOTENPFLCTVVAIIILHYVSKGTEAFTLVEN 2568  
QY 2462 LHLRYALTEYRDVNTGPMRYMLGNGVPAFTIGLAVGLDPBEYGNPDPCMTLSITVTLW 2521  
DB 2569 LHYRYMLTEFRNIDTGMRYHVGWGIPTAVGLADLPDQSGPDMFCLSLQDTLLW 2628  
QY 2522 SFAQPAFAVMSVFLYIIILAAASCAAOQGFKKGVSGIQLQPSFVLLILSLATWLLALL 2581  
DB 2629 SFAQGVGTAVI IINTVLFVLSAKVSCQKHAYERKGVSWLRTAFILLILVTNVLGILL 2688  
QY 2582 SVNSDTLFFYLVATNCICIGPPIFLSYVVLSEVKRALK-LACSRKPSDPALTTKST- 2639  
DB 2689 AVNSDTLSFYLVLAASCLQIGIFVLLPHCVAHREVRKGLAVLAGKTLQDLSATTRATL 2748  
QY 2640 LTSYNSCPSPYADG--RLYQPYGDSAGSLHSTSRSGKSPSYIPLLRSSALNPQGGP 2697  
DB 2749 LITSLSNNTYSEGPDMRLTALGEBSTASLDSTRDEVO-----KLVSQGPARG 2798  
QY 2698 GLADPGSLFL-EGODQOHDPTDSDSLSDQSGSYASTHSSDSEEEHEEEEAAP 2756  
DB 2799 NHGEPDTSFIPRNSKKAHGPDSDDSELSL-DEHSSSYASTHSDSEDDGGAEDK---- 2853  
QY 2757 GEGQMSLPGCARLPLHSTPK-DGPGGCKAPWPD--FGTTAKG----- 2800  
DB 2854 ---WNPAGPA-----HSTPADALANHVPAWMPDSESLAGSDSELDTEPHLYETKV 2903  
QY 2801 ---SSGNAPBEERLRENGALSRGSLGPLPGSSAOP---HKGILKKKCL--PTIS 2848  
DB 2904 SVELHROAGNHCGRSDRSDSEGLAK-----PVAVLSQPOEORCKILNKYTPPELP 2958  
QY 2849 EK--SSILRLPLEQCTGSSRGS-SASEGRSGGPP-----RPPROSLQOQLNGVMP 2898  
DB 2959 EGPLKSLRERKLADCEQSPTSRSTSLGSDGVHATDVCITITKPRRPERHNLNGV--- 3015  
QY 2899 AMSIKATVDEDSGSE 2915  
DB 3016 AMNVRTGSAQANGSDSR 3032

RESULT 9  
US-09-737-149-30  
Sequence 30, Application US/09737149  
Patent No. US2002007466A1  
GENERAL INFORMATION:  
APPLICANT: Spaderna, Steven K.  
APPLICANT: Quinn, Kerry E.  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Muralidhara, Padigaru  
APPLICANT: Spletter, Kimberly A.  
TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-620 CIP  
CURRENT APPLICATION NUMBER: US/09/737,149  
PRIOR APPLICATION NUMBER: 2001-06-15  
PRIOR APPLICATION NUMBER: 60/170,564  
PRIOR FILING DATE: 1999-12-14  
PRIOR APPLICATION NUMBER: 60/173,165  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,362  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: 60/173,544  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 60/174,404  
PRIOR FILING DATE: 2000-01-04  
PRIOR APPLICATION NUMBER: 60/174,962  
PRIOR FILING DATE: 2000-01-07  
PRIOR APPLICATION NUMBER: 60/223,929  
PRIOR FILING DATE: 2000-08-09  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 30  
LENGTH: 3034  
TYPE: PRT  
ORGANISM: Mus musculus

US-09-737-149-30  
Query Match 57.7%; Score 8974.5; DB 9; Length 3034;  
Best Local Similarity 57.3%; Pred. No. 0;  
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;  
1 MSRPATGVLPPEPPLLLLLLLLLLPPPLGQVQPCPSISGRGSGS-----S 48  
DB 124 LRSRARGELRSP-----AVASVQLGDAL--CPRAAGGAASLSVLEALTNPFA 172  
QY 49 GACAPM-----GMLC--PSSASNLMLYTSRCRDAGTELGVPHHDLGVWCPESEA 99  
DB 173 CSCPVAAGTCGRGRPTCLRPFGSALRLVCAIGRAAG-----VNV----- 213  
QY 100 HPLPAPBGCWSCRLGIGIGHLSPOGKLTLPREHPCLKAPRLACSCCKLAQAGLRAG 159  
DB 214 -----ELVIQNTSGTPSSPSV-SFSL--LNLSP--RAG 243  
QY 160 --ERSPEESLGRARRKNVNTAPQOPPSYQATVPENQAGTVPASLRAIDPEGRAGLE 217  
DB 244 VVRSS-----RGTGSSSTSPQPLPSYQVSPENRAGTAVIELRAHDPEDGAGRIS 296  
QY 218 YTMALPDSRNOFPSLDPVTGAVTTABLDREYTSYVFRVTAODHGMRRSALATLT 277  
DB 297 YQMEALPERSNGYFLIDAATGAVTTASLDRETDTHVLKASAVDHGSPRSAAVYLT 356  
QY 278 LVTDPTNDHPVPEQOEYKRSLENNLEGVYVTVARATGDAPPNANILYRLLEGSGSPS 337  
DB 357 TVSDTNDHSFVPEOSEYRERLENLEGVYVTVARATGDAPPNANILYRLLEGSGSG 413  
QY 338 EYFEIDPNSGIVRTGTPVYDREVEVSQTLVEASDGRDPGPRSTAAVFLSEVDNDNAP 397  
DB 414 -VFBIIDASGVVTRAVVDRBEAAYQLLVANDGRNDPGLASATVHIVEDENDVY 472  
QY 398 QSEKRYVQVEDYTPGAPVLRVTAASDRDKSNAVHYSINSNAKRGQFYIDAQTAID 457  
DB 473 QSEKRYVQVEDYVAVTVAVTVATVQATDRQONNAIHYSIVSKQGFYHLSLSGSD 532  
QY 458 VVSPIDYETTKETLRVAADGGPRPLNSGLVYVQVLDINDNAPIFSTPFOATLES 517  
DB 533 VTNPLDFEALIREYTLRIKADGGPRPLNSGLVSVQVLDVNDNAPIFVSSFOAVLEN 592  
QY 518 VPLGYLVHVOAIDADAGNARLEAYRLAGVH-----DPEFTINNGTG 560  
DB 593 VPLGSHVLIHQAVDADAGNARLQYRLVDTASTIYGGSSVDSBNPASAPDFPQJHNSG 652  
QY 561 WTSVAELDREVDYSGVEARHDGTPALTSASVTVYVLDVNDNPPFOPEYTVRN 620  
DB 653 WITVCAELDREVEHYSGVEAVDHGSPAMSSASVITVLDVNDNPPFTVYELRLN 712  
QY 621 EDAAVTSVTVSAVDRDAHSVITYQITSGNTRNFSITSGGGLVSLALPLDYKLERQ 680  
DB 713 EDAAVSSVTLIRADRDANSVITYQLTGANTRNFPALSSGGGLITALLDLKQERO 772  
QY 681 YVLAVTASDGTRODTAQIVNVNTDANTHRPVFOSSHVTVNNEBRPAGTTVVLISATBED 740  
DB 773 YVLAVTASDGTSSHQAQVINTDANTHRPVQSHHYTVSEDRPVGHSIATISATBED 832  
QY 741 TGENNARITFMEDSLPORIADTGAVTQALDVEDQVSYLATAITANGIPKOSDPTY 800  
DB 833 TGENNARITVLPDPPQFRIIDPTGTITMELDVEDQAAVYLAITAOQNGIPQSDTTS 892  
QY 801 LEILVNDVNDNAPQFLRDSYQGVYEDVPFTSVLQISATDRSGLNGVFTYFGGDDG 860  
DB 893 LEILVNDVNDNAPFLRDPFYQGVYEDNAPPSVSVQVSTDRDSGPNGLLTPFGGDDG 952  
QY 861 DGDFTVESTSGIVRTLRKLDRENVAAQYVLRAYAVDKMP-PARTMEVTVYVLDVNDP 919  
DB 953 DGDFTYIEPTSGVIRKQRLDRENVAVVYVLMALAVRGSFNPASVAGVQVSLDINDP 1012  
QY 920 VFEODEPVDVFEVENSPIGLAVARVATPDDEGTNAQIMVQIYEGNIPVYFOLDIFSGLT 979  
DB 1013 VFEODELIFVEENSPVGSVVARIRANDPDEGPNQIITQIYEGNVFVYFOLDILSGDLR 1072



QY 980 ALVLDYEDRPYVYVYIQTASAPVSRATVHVLLDRNDNPVLAGNEFILLFNNTYTNSS 1039  
 Db 1073 ALVLEDPFVBRDYMALVVOATSAPVSRATVHVLLDRNDNPVLAGNEFILLFNNTYTNSS 1132  
 QY 1040 SFFGGAIGRVAPADDPISLSTYSFPERGNELSVLMASNGELKSRALDNNRPLEAMS 1099  
 Db 1133 SFFSGVYIGRIIPADDPISLSTYSFPERGNELSVLMASNGELKSRALDNNRPLEAMS 1192  
 QY 1100 VLVSDBVSHVTAQALAVYIITDEMLTHSITLLEJMSPERFLSPLGLFQAVATLAT 1159  
 Db 1193 VASVDSGHSVTLACTARVYIITDDMLTNSITVRLNMSQKFLSPLSFGVAVLST 1252  
 QY 1160 PPDHVVVNVQNDTAPGHIILNLSVQCPGPGGPPFLPSBDLQEBLYNRSILTAI 1219  
 Db 1253 TTDITFVFNIOQNDTV--SSNIILNVTSALLPGSTRG--RFPSEDIQEOIYNRLTLTYI 1309  
 QY 1220 SAORVLPFDNDI CLRPCENYMRCSVLPFDSAPPLASSVLPRIHVGLRCRCPG 1279  
 Db 1310 SAORVLPFDNDI CLRPCENYMRCSVLPFDSAPPLASSVLPRIHVGLRCRCPG 1369  
 QY 1280 FTGDYCEVEVDLCYRNPCHGRCRREGGYTCLCRDGYTGEHCEVSARSGRCTPGVCN 1339  
 Db 1370 FTGDYCEVEVDLCYRNPCHGRCRREGGYTCECEDEFTGHCQVYNVSGRCAQVCN 1429  
 QY 1340 GGTCTNVLVGGRCDCPSGDFEKPVCQVTTSPHASPITTFGLRORFHTALSPTAK 1399  
 Db 1430 GGTCTNVLVGGRCDCPSGDFEKPVCQVTTSPHASPITTFGLRORFHTALSPTAK 1489  
 QY 1400 RGLLLYNGRFNEKDFVALLEYIOBOVOLTFPSAGSEITTVSPFVPGVSDGQMHVQLKY 1459  
 Db 1490 RGLLLYNGRFNEKDFVALLEYIOBOVOLTFPSAGSEITTVSPFVPGVSDGQMHVQLKY 1549  
 QY 1460 YNKPPLGQTLPOGSEBOKVAVVTVDCDCTVALRFGSVLGNVSCAAGTQGGSKSLDL 1519  
 Db 1550 YNKPPLGQTLPOGSEBOKVAVVTVDCDCTVALRFGSVLGNVSCAAGTQGGSKSLDL 1609  
 QY 1520 TGPPLLVGVYVPLPESFPVVRMROFVGCNRLQVDSRHIDMADPIANNGTVPGCAKNVCD 1579  
 Db 1610 TGPPLLVGVYVPLPESFPVVRMROFVGCNRLQVDSRHIDMADPIANNGTVPGCAKNVCD 1669  
 QY 1580 SNTCHNGGTCTVNMOWDAFSCCEPLGFGGKSCAQEMANPOHFLGSSIVAMHGLSLPISOPMY 1639  
 Db 1670 SNTCHNGGTCTVNMOWDAFSCCEPLGFGGKSCAQEMANPOHFLGSSIVAMHGLSLPISOPMY 1729  
 QY 1640 ISLMERTROADGVLLQATITRGRSITTLQUREGHVMSYEGTGLQASLRLERGRANDGM 1699  
 Db 1730 ISLMERTROADGVLLQATITRGRSITTLQUREGHVMSYEGTGLQASLRLERGRANDGM 1789  
 QY 1700 HHAQLALGASGPGH---AIIISPDYGOQRAEAGNLGPRHLGHLNITVGGI PGAPAGV 1754  
 Db 1790 HHAQLALGASGPGH---AIIISPDYGOQRAEAGNLGPRHLGHLNITVGGI PGAPAGV 1848  
 QY 1755 ARGFGGICGVVSPDPBGVNSLDPHSGHSINVEGCSLPDPBCDNPCCANSYCSNMDS 1814  
 Db 1849 ARGFGGICGVVSPDPBGVNSLDPHSGHSINVEGCSLPDPBCDNPCCANSYCSNMDS 1908  
 QY 1815 YSCSCDPGYGDNCTNVCDLNFCEHOSVCTRRKPSAPHGYTCECPNVLGYPCTETRIDPC 1874  
 Db 1909 YSCSCDPGYGDNCTNVCDLNFCEHOSVCTRRKPSAPHGYTCECPNVLGYPCTETRIDPC 1966  
 QY 1875 PRGMWGHPTGFCPCNDVSKGPPDCNKTSGEHCCKENHTRPGSPPTCLCDGYPTGSLSR 1934  
 Db 1969 PRGMWGHPTGFCPCNDVSKGPPDCNKTSGEHCCKENHTRPGSPPTCLCDGYPTGSLSR 2028  
 QY 1935 VCDPBGDGPCCARGLVIGRCDCRCDNPPAVVTNGCEVNVYDSCRAIIBAGIMWRTRFGLP 1994  
 Db 2029 VCDPBGDGPCCARGLVIGRCDCRCDNPPAVVTNGCEVNVYDSCRAIIBAGIMWRTRFGLP 2088  
 QY 1995 AAAPCPKSGFETAVRHCDEHRGMLPENTFNCTSIYFSLKGPABRIQUNESGLDSGRSQ 2054  
 Db 2089 AAAPCPKSGFETAVRHCDEHRGMLPENTFNCTSIYFSLKGPABRIQUNESGLDSGRSQ 2148

QY 2055 IALLIRNATORTAGYEGSDVAVAYQATRLLAHSTORCGFGLSANTQDVHFTENILRVGSA 2114  
 Db 2149 IAKALRNATORTAGYEGSDVAVAYQATRLLAHSTORCGFGLSANTQDVHFTENILRVGSA 2208  
 QY 2115 ILDTANKHMEILQOTBEGTAMLLQHYEAYASALAAONRHLYLSPTVTPNIVSVRL 2174  
 Db 2209 ILAPTEASWEQIOESKGAQQLNHFPAYSNNVARNKRYTLRPFVITVAMMILAVIIF 2268  
 QY 2175 DKGNFAGAKLPYREALRGEPDPLETTYILPESVY---RETPVVR-----PAGE 2221  
 Db 2269 DKGNFAGAKLPYREALRGEPDPLETTYILPESVY---RETPVVR-----PAGE 2328  
 QY 2222 GEAQPEELARORRHPELQGEAVASVYIYRTLAGLLPHNYDPDKSLAVPKRPIINTP 2281  
 Db 2329 GEAQPEELARORRHPELQGEAVASVYIYRTLAGLLPHNYDPDKSLAVPKRPIINTP 2388  
 QY 2282 VVSIIVHDBELRALDKPTVQFLLLETERKPCIVFNNHSLVSGTGKMSARGCEV 2341  
 Db 2389 VVSIIVHDBELRALDKPTVQFLLLETERKPCIVFNNHSLVSGTGKMSARGCEV 2448  
 QY 2342 VFRNESHVSCCNHMTSPAVLMDVSRRENGEILPLKTLTYVALGVTALALITPFLTL 2401  
 Db 2449 VFRNESHVSCCNHMTSPAVLMDVSRRENGEILPLKTLTYVALGVTALALITPFLTL 2508  
 QY 2402 RILSNHOGIRRNITPAIQLAQLVFLGINDADLPACTVIAIILHFLYLCFPMALIA 2461  
 Db 2509 RILSNHOGIRRNITPAIQLAQLVFLGINDADLPACTVIAIILHFLYLCFPMALIA 2568  
 QY 2462 LHLVRLTEVDVNTGPRPFYVYMGVPAITGLAVGLDEGVGNPFCULSTYDTLIW 2521  
 Db 2569 LHLVRLTEVDVNTGPRPFYVYMGVPAITGLAVGLDEGVGNPFCULSTYDTLIW 2628  
 QY 2522 SFGAVFAVAVSMYVLYIARASCAAROGFEKKGPVSGQSPFVAVLYLSATMTLAL 2581  
 Db 2629 SFGAVFAVAVSMYVLYIARASCAAROGFEKKGPVSGQSPFVAVLYLSATMTLAL 2688  
 QY 2582 SVNSDTLIFHYLPATNCIOGPFILSVVLSKEVRALK-LACSRKSPDPALTYKST- 2639  
 Db 2689 SVNSDTLIFHYLPATNCIOGPFILSVVLSKEVRALK-LACSRKSPDPALTYKST- 2748  
 QY 2640 LTSYVNCPSYADG--RLVQYFGDSAGSIHSTSGKSGQSPYIIFPLREBALNPGQPP 2697  
 Db 2749 LTSYVNCPSYADG--RLVQYFGDSAGSIHSTSGKSGQSPYIIFPLREBALNPGQPP 2798  
 QY 2698 GLADPGSLFL-EGDOQDHPDVTDSDSLSEDDGSGVASTHSDSEEBEEREEKAAP 2756  
 Db 2799 GLADPGSLFL-EGDOQDHPDVTDSDSLSEDDGSGVASTHSDSEEBEEREEKAAP 2853  
 QY 2757 GEQGWDSLPGFABRLPLHSTPK-DGGRPGKAPWCD--FGTTAKG----- 2800  
 Db 2854 GEQGWDSLPGFABRLPLHSTPK-DGGRPGKAPWCD--FGTTAKG----- 2903  
 QY 2801 -----SSNGAPAEERLREKGDLSREGSLGRLPGSSAOP---HKGLKKCC--PTIS 2848  
 Db 2904 -----SSNGAPAEERLREKGDLSREGSLGRLPGSSAOP---HKGLKKCC--PTIS 2958  
 QY 2849 EK--SSILRLPLBOCTSSRGS--SASRSGSGAPP-----RPPROSQLOQLANVPI 2898  
 Db 2959 EK--SSILRLPLBOCTSSRGS--SASRSGSGAPP-----RPPROSQLOQLANVPI 3015  
 QY 2899 AMSTIKAGTVDEDSGSR 2915  
 Db 3016 AMSTIKAGTVDEDSGSR 3032

RESULT 10  
 US-10-131-409-70  
 ; Sequence 70, Application US/10131409  
 ; Publication No. US20030199465A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Malayankar et al.  
 ; TITLE OF INVENTION: No. US20030199465A1el Polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 15966-675C1P1CONT1

```

CURRENT APPLICATION NUMBER: US/10/131,409
CURRENT FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: 09/898,954
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: 60/182,733
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 60/182,724
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 60/183,896
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: 60/184,497
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/224,157
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 60/184,482
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/184,744
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: 60/197,083
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/233,405
PRIOR FILING DATE: 2000-09-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 135
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 3034
TYPE: PR
ORGANISM: Mus musculus
US-10-131-409-70

```

```

Query Match 57.7%; Score 8974.5; DB 12; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

```

```

1 MRSPTATGVLPTPPPLLLLLLPPLLDGVOGCRSLGSRGSG-----S 48
124 LRSASGAEIRSP-----AVRSVPGIDAL--CFPAAGGGAALSTSVLRATNTPA 172
49 GACAPM-----GMLC--PSSASNLMLYTRCRDAGTELTHLVPHHDLRWCPESBA 99
173 GSCPPVAGTCRCRGPICLRPGSHELRLVCLGRABA-----VWV----- 213
100 HILPPAPBCPCWSCRLIGIGHLSPQKLTLPBHPCLAPRLRCSCQAQAPGLRAG 159
214 -----ELVIGATSGTSPSSPSV--SPGL-----LNLSPG--RAG 243
160 --ERSPEBSLGGRRKRVNTAPQOPPSYQATYPENOPACTPVASLRALDPDEGAGRL 217
244 VVRS-----RRGTGSSSTSPQPLPSYQVSVENEPACTAVITELRAHDPDGDGRLS 296
218 YTMADLPDSRNPQFSLDPYGAVTTAELDRKTSKTHVFRVTAODHGMRRSALATLT 277
297 YQMBALPDSRNGFLIDAATGAVTTARSLDRKTHLVKSAVDHSPRRSAATLTLY 356
278 LVTDTMDHDPVROEYKSLRENTLEVGEVLTVRATDGPAPNANILYRLLSGGSGS 337
357 TVSDTNDHSPVFRQSEYRERIRENTLEVGEVLTFRATDGPANANRXYRLLSGAGG--- 413
338 EVFRIDRSGVITRGVNDREBVSQLYTBASDQGDPPRSTTAAVLSVEDDNDNA 397
414 -VEIDRSGVFRVRAVVDREBAEYQLLVANDQGNPGLSASAVHIVDEBNDNP 472
398 QFSEKRVVQVREDVTPGAPVLRTVTSADRDGNSNAVHYISMGNGAGCFYLDAGTALD 457
473 QFSEKRVVQVREDVTPGAPVLRTVTSADRDGNSNAVHYISMGNGAGCFYLDAGTALD 532
458 VVSLDYEYTKETTLRVADQGRPLSNVSGVTVVTDINDNAPITVSTPQATVLES 517
533 VINPLDPEARBYRLIKADQGRPLINSGLVSVQVLDVNDNAPITFVSPFOAAVLN 592
518 VPLGVLVYQALDADGDNARLELYLAGVH-----DFTTNGTG 560
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

```

593 VELGSHVLIHQAVDADAGENARLQYRLVDTASTIVGSSVDSNPASAPDFPQIHNSSG 652
561 WTSVAALDRREVDYSGVEARHDGTALTSASVSTVVDVNNNPFPTQPEYTVLN 620
653 WTVCAELDRREVEHYSGVEAVDHGSPMSSASVITVLDVNDNPFTVPVEELRLN 712
621 EDAAVGTSVTVASADRDASHVITYQTSGNTRNFSITSGGGLVSLALPLDYKLEHQ 680
713 EDAAVGSSTVLRADRDANSTVITYQTSGNTRNFPALSSGGGLTALPLDYKLEHQ 772
681 YLAVTASDGTRODTAQVNVNTDANTRPVFOSSHVTYVNNEDRPACTTVVLSATDDE 740
773 YLAVTASDGTSHTAQVFINVTDANTRPVFOSSHVTYVSEDRPVGTSIATISATDDE 832
741 TGNARITYEMEDSLPDRIDADTGAVTQAEVDYEDVSYTLATDANGIPQSDPTTY 800
833 TGNARITYEMEDSLPDRIDADTGAVTQAEVDYEDVSYTLATDANGIPQSDPTTY 892
801 LEILVNDVNDNAPQELRDSYQGSVYEDVPPTSVYQISATDRDGLNGRFTYPCGDDG 860
893 LEILVNDVNDNAPRRLRDPYQGSVYEDVPPTSVYQISATDRDGLNGRFTYPCGDDG 952
861 DDDPFTVESTGTGVTLRRLDRNVAAQVLRAYAVDKMP-PARTMEVTVTVLVNDNPP 919
953 DDDPFTVESTGTGVTLRRLDRNVAAQVLRAYAVDKMP-PARTMEVTVTVLVNDNPP 1012
920 VEEODEPQVFEVENSPIGLAVAVTATPDDEGTNAQVQIYEGNIPEVFDIDISELT 979
1013 VEKDELEFVEBNSPVGSVVARIRAMPDRBSPNQQIITYQVAGVPEVFDIDISELT 1072
980 ALVDDYEDRPEYVTVIQAATSAVLVSRAVTVHRLDRNDNPPVNLFPNNYTNNS 1039
1073 ALVEIDFEBRDPYMLVQATSAFLVSRAVTVHRLDRNDNPPVNLFPNNYTNNS 1132
1040 SPPGALGRVPAHDIDISLTYSPERGENSLVILMASTGELKLSRALDNRPEALMS 1099
1133 SPPSGVIGIRPAHDIDISLTYSPERGENSLVILMASTGELKLSRALDNRPEALMS 1192
1100 VLVSDGVHSTVQALRTYITIDEMLTHSITRLTLDMSPERLSPILGLFQVAVATLAT 1159
1193 VSVSDGHSVTVLCTLRATYITIDMLTNSITVRLNMSQKLSLSTFVSGVATVYST 1252
1160 PPDHVVVFNVDRTDAPGHHILVSLVGOQPPGSGPPLESEDLQERLYNLSLTLAI 1219
1253 TKDDIFVFNQNDTVV-SSNIIWTFSSALLPQGTG--RFPSESDQEBQIYNRLTLTYI 1309
1220 SAORVLPDPDNTCLABPCENYKCVSLRPSASAFIASSSTLPRPIHVGLRRCRCPG 1279
1310 SAORVLPDPDNTCLABPCENYKCVSLRPSASAFIASSSTLPRPIHVGLRRCRCPG 1369
1280 FTGDYCETEVDLCYSPGCPGHRCSRREGYVCLCRDGYTGEGHCEVSARSAGRCTGVCKN 1339
1370 FTGDYCETEVDLCYSPGCPGHRCSRREGYVCLCRDGYTGEGHCEVSARSAGRCTGVCKN 1429
1340 GGTGVNLLVGAFKCDPESGDFEKPVCQVTTNSFPASHPTTFEGLRORHFTLALSPATKE 1399
1430 GGTGVNLLVGAFKCDPESGDFEKPVCQVTTNSFPASHPTTFEGLRORHFTLALSPATKE 1489
1400 RGLLLYNGRFBKBDPVALVLEBOVLTFSAGSSTTVSPFVGVGSDGQMHVTVOLKY 1459
1490 RGLLLYNGRFBKBDPVALVLEBOVLTFSAGSSTTVSPFVGVGSDGQMHVTVOLKY 1549
1460 YNKPPLGOTGLPOGSPBOKVAVVTVYDGDVVALRFGSVLANSYCAAOGTGSGSKSLDL 1519
1550 YNKPPLGOTGLPOGSPBOKVAVVTVYDGDVVALRFGSVLANSYCAAOGTGSGSKSLDL 1609
1520 TGPILLGVPPLPESFPARMQPVGCNRLQVDSHIMADITANGVTPGCPAKENVD 1579
1610 TGPILLGVPPLPESFPARMQPVGCNRLQVDSHIMADITANGVTPGCPAKENVD 1669
1580 SNTCHNGGTCVQNMQAFSCCEPLGFGKSCAQMNPQHFLLSSIVAMHGLSLPTISQPPY 1639
1670 GTSQNGGTCVQNMQAFSCCEPLGFGKSCAQMNPQHFLLSSIVAMHGLSLPTISQPPY 1729

```

```

QY 1640 LSLMPTROADVLLQAITRGSTITLQUREGHTMLSVETGLOASLRLBGRANDGM 1699
DB 1730 LGIMFTREDEGVLMETAGTSSRLHLQILNSTREREVSIGPSDVAWSQLSKSRITDGM 1789
QY 1700 HHAOLALGASGGRH-----ALISPDYGOORABGNLGPRLHGLHLSNITVGGIPBAGV 1754
DB 1790 HHLLEL-RSABEGKDIKTLAVWTLVDYGMDOSTVOQNLPGIKKRTIYIGVETEDKVS 1848
QY 1755 ARGFRCLOGVAVSDTPREGVNSLDPESHGSIINVEQCSLPDPCDSNCPANGYCSNDMS 1814
DB 1849 RHGFPGCNGVAMGERTSTMIATLNMMDALKRVKOCGVEDVEDCASPCHPAPPCADTMS 1908
QY 1815 YGSCDPCGYGNCNTVCDLNPCEHOSVCTRRKPSABHYTCBCPPNYLGPYCETRIDPC 1874
DB 1909 YSCIDRGYFGKKCVADACLLNCPCKHVAACVRSBPNTPRGYSCEGPHGYOCENKVDJPC 1968
QY 1875 PRGMWGHPTGCPGNCNDVSGKPDNDCKTSGECHKENHRRPFGSPCLLDCCYFGSLR 1934
DB 1969 PRGMWGNPVGCPCHCAVSGQSPDCKNTNGCQCCKENYKPPADACLPDCCDFPFGSHR 2028
QY 1935 VCDPEGDQCPCKRGVIGRQCDRCNDPFAVTNGCEVNDSCPRALIEAGIMPTRTFGLP 1994
DB 2029 ACDMDTGQACCKRGVIGRQCDRCNDPFAVTNGCEVINGCPRAEAGIMPTRTFGLP 2088
QY 1995 AAAPCPKSGFTAVHCDERHGMPLPMLFNCTSTIFSEILKGAERLQENESGLDSGRSQ 2054
DB 2089 AAAPCPKSGVGNVHCHSGEKGMPLPELFNCTSGFVLDKALNEKLNREHMDGNRSLR 2148
QY 2055 LALLRNNTQHTAGYFGSDVYKAYXOLATPLLAHESQRFGLSATQDVHTENLARGSA 2114
DB 2149 LAKALRNNTQNSTLFGNDVRYAYOLLAIILOHESRQOFDLAAREANFHDVHTGSA 2208
QY 2115 LLDLTANKHMBELIOTEGGTAWLLOHYEAYASALAQNMSHTLSPPTIYTPNIVISVRL 2174
DB 2209 LLAAPTASWEOIQSSEGAQOLHNFEPFNSVANRKTILRFVITVANMILAVIDIF 2268
QY 2175 DKGNFAGAKLPRYELRGHQPDLFTVLPESVF-----RTPPYVR-----PAGP 2221
DB 2269 DKLNTGQVPRFEDIOBELPRELESVSFPADTFKPREKKGPRVLRNRRRTBLTAQP 2328
QY 2222 GEAPREBELARORHPELSGSEAVASVIYRTLAGLLPHNDPKRSIRVKTINTP 2281
DB 2329 EPPARETSSSRHRHPPGPPAVALVITYTLQOLPEHDPHRSILRNRRVINTP 2388
QY 2282 VVSIHVDDEBLIPALDKPVTVQORLLETERTKPICVFNMHSILVSGTSGMSARGEV 2341
DB 2389 VVSANVYSEGTLPESLQRPILVERSILETTERSVCVFMHSLDTGCTGMSAKGCEL 2448
QY 2342 VFRNESHVSCQCNMTSPAVLMDVSRRENGETLPLKTLTYVALGVTLAALLTFPFTLL 2401
DB 2449 LSRNTHVTCQCSHSCAVLMDISRREHGEVLPKITYVALSLVALVAFTLLSLV 2508
QY 2402 RLRSNOHGIRNRLAALGLAVFLGIGNADLPACVIVAIIHLPLXLCFSPALLIEA 2461
DB 2509 RLRSNLSIHKNLLAALPFSOLIMVGNQTEHPCLCVAAIIHLHYVSMGFPAATLVEN 2568
QY 2462 LHLVRLTEVRDVNTGPMRFYMLGMVPAFTTGLAVGLDPGNGPFCMLSTYDTIM 2521
DB 2569 LHVYMLTEVNIDIDGPMRFYHVGMPALVYTGAVGLDPOGNGPFCMLSTYDTIM 2628
QY 2522 SPAGVAVAVMSVFLYILABASCAQORGFEXKGPVSGLOPSAVILLLSATWLLAL 2581
DB 2629 SPAGVGAVIILINVIYVLSAKVSCORKHYYERKGVVSMRTAFLLILLYATWLLGL 2688
QY 2582 SVNSDPLLHLYLFAONCIGPFIPLSYVLSKVRKALK-LACRKSXPALATKST- 2639
DB 2689 ANSDTLLSFHYLPAFSCLOGLIFVLFHCVAREVRKHLRAVLAGKQLDLSATRTYL 2748
QY 2640 LTSYNCSPPYADG--RLYQPYGDSAGSLHSTRSGKSPSYIPLLEBSALNFGQPP 2697
DB 2749 LTRSLNKNNTYSEBGMRLTALGESSTALSDSTTRBGOV-----KLSVSSGPARG 2798

```

```

QY 2698 GLGDGSLFTL-EGDQDQHPDPTSDSLSLBDQSGSYASTHSDSEEBEEREAAP 2756
DB 2799 NHGEPDTSFIRPNSSKKAIGPDSDSLSL-DEHSSYASSHTSPSDDGGAEDK----- 2853
QY 2757 GEQGWDSLGRGAEPLPHSTPK-DGGRPGKAPPGD--FETTKX----- 2800
DB 2854 ----WNPAGGPA-----HSTPKADALANHVAGPDSLSLSDSEEDTEPHLKEVKV 2903
QY 2801 ----SSGNAPERLRENGDALSRGSLGPLPSSAOP-----HGILKKKCL--PTIS 2848
DB 2904 SVELHROAQHCHGDRPSPDSSGLAK-----PAVAVSSQFOEQRKGLIKKVTYPPPLP 2958
QY 2849 EK--SSILRLPLEOCTGSSRGS-SASGSRGAPP-----RPPRQSLQOLNGVPI 2898
DB 2959 EQLPSRLREKLADCBGPTSSRTSSLSGSGVATDCVITIKTPRREPRGHEHLNGV--- 3015
QY 2899 AMSIKAGVDEBSSGSE 2915
DB 3016 AMNVRTGSAQANGSDSE 3032

```

```

RESULT 11
US-10-120-801-52
; Sequence 52, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Foad
; APPLICANT: Topper, James N.
; APPLICANT: Malysankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/286334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: mouse
US-10-120-801-52

Query Match 57.7%; Score 8974.5; DB 12; Length 3034;
Best Local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

1 MRSPATGVPLTPPPPLLLLLLLLLLPPPLGDPGCPKSLGSRGSG-----S 48
DB 124 LKRSARGAELRSP-----NVRSVFGLGDLN--CPAAGGASLSTVLEAITNPPA 172

```

QY 49 GACAPM-----GMLC--PSSASNLMLYTSRCRDAGTELTHLVPHHDLRWCEBSA 99  
 Db 173 CCGPFAVGCGCRGPICLAPGSSAEHLVCLALGAAAGA-----VWV----- 213  
 QY 100 HILPLPAPBGCPMSCLLIGIHLSPQKLTLPBEPCLKAPRLRCSCKLAQAPGLRAG 159  
 Db 214 -----ELVIAQTSGPSESPSV-SPSL-----INTLSQP---RAG 243  
 QY 160 --BRSPEBSIGGRKKNVNTAPOPQPSYQATYPENQAPGTVAASLRALDDPDEGEGGLE 217  
 Db 244 VVARS-----RRGSSSTSPPQPLPSQVSVBENEPAGTAVIELRADPDEGDGRIS 296  
 QY 218 YTMADLPDSRNOFSLDPVTGAVTAAEBLDRKTSHYVRVYAOHGMPPRSALATLT 277  
 Db 297 YQMBALPDRSNQYFLIDATGAVTTAKSLDRKTKOTHLKVASVDHSGPRSAATYLLV 356  
 QY 278 LVTYNDHDPVFPQOEKESLRLENLVGEVLTVRAITDGPANANILYRLLEGSGSPS 337  
 Db 357 TVSDTNDHSPVFPQSEYRERIRENLVEYEVLTIRATDGPANANIRYLLLEGAGG--- 413  
 QY 338 EYVEIDPBRSGVITRGPVDRBEVESYQLTVEASDQGRDPGRSTTAAVLSVEDDNDANP 397  
 Db 414 -VEIDARSGVATRAVVDREBAEYQLVLEANDQGNPGLSASATVHLVDEENDNTP 472  
 QY 398 OFEKEKYVVOVREDVTPGAPVLTAVTASDRDKSNVAVHSIMSGNARGOPFLDAOTGALD 457  
 Db 473 QPBEKRYVVOVPDVAVNTAVTLVQATDRQGNMAIHISIVSGNLKQPYLHLSGSLD 532  
 QY 458 VVSPLDYETTKETYLVRADQGRPPLSNVSLVTVQVLDINDNAPL FVSTPQATVLES 517  
 Db 533 VINPLDFEALREYTLRIKAQDGRPPLINSGLVSVQVLDVNDNAPL FVSSPQAVLEN 592  
 QY 518 VPLGYLVTHVQALDADAGDNARLELYLAGVH-----DREPTINNGG 560  
 Db 593 VPLGSHVLTQAVDADAGENARLQYRLVDTASTTVGSSVDSNPNASADPFPQIHNSG 652  
 QY 561 WISVAALDRREEDPYFSFGVEARHDGTPALTASASVTVLDVNDNPPFTQCEYTVRLN 620  
 Db 653 WIVVCAELDRREVEHISFGVEAVDHGSPANSSASVITVLDVNDNPPMTQVYELRLN 712  
 QY 621 EDAVAGTVTVTSAVDRDAHSVITTOITGNTNRPESITQSOGGVLALPLDYLERQ 660  
 Db 713 EDAVAGSSVTLTARBRDANSVITTOITGNTNRP FALSOGGGLITLALPLDYQERQ 772  
 QY 681 YVLAVTASDGTRODTQIVVNTDANTHRPVFQSSHVTAVVNBDRPAGTVVLISATDEP 740  
 Db 773 YVLAVTASDGTSHTAQVFINVTDANTHRPVFQSHVTASVSEDRPVGISITATISATDEP 832  
 QY 741 TGENARITVEMEDSIPOFRIDADTGAVTTOAELEDEQVSYTLATARDNGIFOKSDTTY 800  
 Db 833 TGENARITVLEBVPQFRIDPDTGITYTMBELDYEDQAAVTLATITQDNGIFOKSDTTS 892  
 QY 801 LELLVNDVNDNAPQFLADSIQGSYTEDVPPTSYLQISATDRQSGLANGVFTTFQGGDQ 860  
 Db 893 LELLVNDVNDNAPFLADSIQGSYVEDAPPTSYLQISATDRQSGPGRLLYTFQGGDDG 952  
 QY 861 DGDPIVESTGIWTLRLRRENOAVVLAAYAVDKAMP--PARTPMEVTVTVLDVNDNP 919  
 Db 953 DGDPIESTGIVITQRLDRRENOAVVNLMAVDRSPNPLASVGIQVSLVDINDNP 1012  
 QY 920 VFBODEFVIVENSPIGLAVARVTATDPDEGTNAQIMYOIVBGNLPEVQOLDIFSGELT 979  
 Db 1013 VFBODELELFVEERSPVGSVVARIRANDPDEGPAQIITQIVBGNVPEVQOLDISGLDR 1072  
 QY 980 ALVLDVDEDEPEVTVLQATSAPLVSRATVHVLDDNDNPPVLGNRELLPNNYVTRRS 1039  
 Db 1073 ALVBLDEVRDNYLVVQATSAPLVSRATVHILDDNDNPPVLPDQIILPNNVYTKSN 1132  
 QY 1040 SFPQGAIGRVPAHPDLSITLTSFPERGNEISLVLNASTGELLSBALNNPPLAAMS 1099  
 Db 1133 SFPQGVIGRIPAHPDLSITLNTFTLQGNELSLILLDPATGELQSLSDLDNNRLEALME 1192

QY 1100 VLVSQVHVTQAOCALRVITITDEMLTSTITRLLEDMSPERFLSPILGLOAVATLAT 1159  
 Db 1193 VVSQSHISVTRALCTLRATITITDDMLTNSITVRLNMOEKRLSLSLFVGVATVLT 1252  
 QY 1160 PPDHVVNVQORDTAPGHHILNLSVQGPFGGPPPLPSBDLQERLYNRSLLTAI 1219  
 Db 1253 TKDDIFVFNQINDTV--SSNITLVFSLALPQTRG--RFPSSDLEQOYINRLLTITI 1309  
 QY 1220 SAORVLPDDNCLAREPENTWCVSLRFPSSAPFLASSVTLFRHIPVGLRRCRPG 1279  
 Db 1310 SAORVLPDDNCLAREPENTWCVSLRFPSSAPFLASSVTLFRHIPVGLRRCRPG 1369  
 QY 1280 FTGDYCETEVLDIYSRPGCPHRCRSREGYTCICRDEYTGHEHCVSARSAGCTGVCKN 1339  
 Db 1370 FTGDYCETEVLDIYSRPGCPHRCRSREGYTCICRDEYTGHEHCVSARSAGCTGVCKN 1429  
 QY 1340 GGTGVNLLVGGFKCDPSPGDEPKPYQVYTNSSFPHASFTTFRGLRQRFHTALSPATKE 1399  
 Db 1430 GGTGVNLLVGGFKCDPSPGDEPKPYQVYTNSSFPHASFTTFRGLRQRFHTALSPATKE 1489  
 QY 1400 RDGILLYNGRFBKEDPVALFIOBOVLTFSAGSSTTVSPFVGVGVSDGQMTVQKX 1459  
 Db 1490 RNALLVNGRFBKEDPVALFIOBOVLTFSAGSSTTVSPFVGVGVSDGQMTVQKX 1549  
 QY 1460 YNKPLLQGTGLPQGBEOKVAVTVVDCDVTGVALRFGSLVAGYSCAAGTQSGSKSLDL 1519  
 Db 1550 YNKPNIGHLGLPHGSGKRVAVTVVDDCDAAVAHFGSVYGVYSCAAGTQSGSKSLDL 1609  
 QY 1520 TGPILLGGVPLDPEFPVPRMROFVCGMRLOVDSHIMADFIANGVYPCGPAKAVCD 1579  
 Db 1610 TGPILLGGVPLDPEFPVPRMROFVCGMRLOVDSHIMADFIANGVYPCGPAKAVCD 1669  
 QY 1580 SNTCHNGGTGVNQMAFSCBCEPLGFGKSCAOGEMANPOHFLSSLVAMHGLSLPTISQPY 1639  
 Db 1670 GTSQNGGTGVNQMAFSCBCEPLGFGKSCAOGEMANPOHFLSSLVAMHGLSLPTISQPY 1729  
 QY 1640 LSLMFRTRQADGVLLQAITRGRSTITLQIREGVMVSEGTGLQASSLRLEBRANDGDW 1699  
 Db 1730 LSLMFRTRQADGVLLQAITRGRSTITLQIREGVMVSEGTGLQASSLRLEBRANDGDW 1789  
 QY 1700 HHAQALASGSGR-----ALSPDYGOQARBNGLGRIRHGLHLSNITVGGIIPBAGV 1754  
 Db 1790 HHAQALASGSGR-----ALSPDYGOQARBNGLGRIRHGLHLSNITVGGIIPBAGV 1848  
 QY 1755 ARGFRGCLQGVVSTPREGVNSLDRSHGESINVEQCSLPDCCDNPCPANSYCSNDMS 1814  
 Db 1849 RHGFRGCMQGVVSTPREGVNSLDRSHGESINVEQCSLPDCCDNPCPANSYCSNDMS 1908  
 QY 1815 YSCSDPGYGDNCNVCDLNFCEHQSVCTRRSPAPHGTYCECPNVLGAPYCETRLIDPC 1874  
 Db 1909 YSCSDPGYGDNCNVCDLNFCEHQSVCTRRSPAPHGTYCECPNVLGAPYCETRLIDPC 1968  
 QY 1875 PRGMGHPRCPCNCDVSGRPPDCNKTSGBCHEKNHRRPBGSPPTCLDCYCPYGSLSR 1934  
 Db 1969 PRGMGHPRCPCNCDVSGRPPDCNKTSGBCHEKNHRRPBGSPPTCLDCYCPYGSLSR 2028  
 QY 1935 VCDPEDEQCPCKRPGVIGRQCDRCNPNFAVTTNNGEVNYSCEPRAIEAGIWWPRTFGLP 1994  
 Db 2029 ACMDMVGQACKRPVIGRQCDRCNPNFAVTTNNGEVNYSCEPRAIEAGIWWPRTFGLP 2088  
 QY 1995 AAAPCKGSGFTAVRHCDEHGMPLPNLFCNTSITPSEBLKFAERLQNNBSGLDSGRSQ 2054  
 Db 2089 AAAPCKGSGFTAVRHCDEHGMPLPNLFCNTSITPSEBLKFAERLQNNBSGLDSGRSQ 2148  
 QY 2055 LALLLNATQHTAGVFGSDVKAAYOLATLHLBSFQSGFGLSATODVHFTENLILVGA 2114  
 Db 2149 LALLLNATQHTAGVFGSDVKAAYOLATLHLBSFQSGFGLSATODVHFTENLILVGA 2208  
 QY 2115 LLDFTANKRMBELIQTEGTAWLLOHYEAYASALQNMHTYLSPTIYTPNIVISVBL 2174  
 Db 2209 LLDFTANKRMBELIQTEGTAWLLOHYEAYASALQNMHTYLSPTIYTPNIVISVBL 2268  
 QY 2175 DKGNFAGALPRYEAALRGQPPDLERTVILPESVF-----RETPPVVR-----PACP 2221

```

Db      2269  DKLNTGQVFPFEBIOEELPRELESSVFPADTFEPPEKKGFPVRLNRRATPLTQAP 2328
Qy      2222  GEAQPEELARORRHPELSQGBAASVIVYRTLAGLLPHNDPDKRSIAVRKPIINTP 2281
Db      2329  EPRARETSSSRKRRHPDEGQFPAALVIVITLQOLPEHDPHRSRLRNREIVIMP 2388
Qy      2282  VVSIIVHDEBELPPALDKPVTVQFRLLETERKPICVFNNHSILVSGTGMARGCEV 2341
Db      2389  VVSANVYSEGTFPLPSLORPLIVERFLETERKSPVCVFNNHSIDTGTGMSAKGCEL 2448
Qy      2342  VFRNESHVSCQCNMTSPVLMVDSRRNGEELPLKTLTYVALGVTALALLTFFPFLT 2401
Db      2449  LSRNTHVTCQCSHSSASCAMIDISRBRGEVLPLKITVYALSLVALVAFVLSIV 2508
Qy      2402  RLRNNOGIRNLTALGLAQVFLGNOADLPACVIAIILHPLTCFSPALLA 2461
Db      2509  RLRNNSHSHKMLALPFSOLTFWVGINQTEPPLCTVAILHYVSMGFPAATLVEN 2568
Qy      2462  LHLVRLTEVDVNTGPMKFTYMLGMVPAFTGLAVGLDPEGYGNPFCMLSIYDTLIM 2521
Db      2569  LHVYMLTEVRNIDGPMKFTYVGMGIPALVITGLAVGLDPOGYGNPFCMLSIYDTLIM 2628
Qy      2522  SPAGVAAVAKSVLYTLAARASCAQROGFEKKGPVSGLOPSAVILLISATVLLAL 2581
Db      2629  SPAGVGVIIIVITVYLSAKVSCQKHHYERKGVSMRTAFILLLVATVLLGL 2688
Qy      2582  SVNSDTLLFHYLPATCNICIGPEIFLASYVLSKVRKALK-LACRKRSPDPAITTKST- 2639
Db      2689  AVNSDTLLFHYLPATCNICIGPEIFLASYVLSKVRKALK-LACRKRSPDPAITTKST- 2748
Qy      2640  LTVSYNCPYADG--RLYOPYDGSAGSLHSTSRSGKSPSYIPLLEESALNPGQGP 2697
Db      2749  LTVSYNCPYADG--RLYOPYDGSAGSLHSTSRSGKSPSYIPLLEESALNPGQGP 2798
Qy      2698  GLGDPGSLFL-EGDQDQDPTDSDSLLEDQSGVASTHSSDSEEEBEEBAAP 2756
Db      2799  NHGEEDTSFIPRNSKKAHGPDSDSSELSL-DEHSSYASSHTSSEDDGGEABDK 2853
Qy      2757  GEGQWDSLILGQALRLPLHSTPK-DGPGPGKAPPGD--FGTTAKE----- 2800
Db      2854  WNPAGGA-----HSTPKADALANHPAGHFDSESLAGSDSEEDLDETERPHLKVEYKV 2903
Qy      2801  SSGNGAPEERLRENGDALSRGSLGPLPGSSAOP--HKGLKKKCL--PTIS 2848
Db      2904  SVELHQQGNHCGRPFSDPESGVLAQ--PVAVASQPOBOKGILKKKVTYPPPLP 2958
Qy      2849  EK--SSLLRLPLEQGTSSRGS-SASESGSGGPPP-----RPPROSLQEOUNGWPI 2898
Db      2959  EQPLKSRLEKLADEOSPTSSKRTSLSGDGVNATDCVITIKTPRPREGREHLNGV--- 3015
Qy      2899  AMSIKAGTVDEDSGSE 2915
Db      3016  AMNVRTGSAQANGSDSE 3032

```

RESULT 12  
US-10-150-811-70

```

; Sequence 70, Application US/10150811
; Publication No. US20040010120A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar et al.
; TITLE OF INVENTION: No. US20040010120A1 Polypeptides and Nucleic Acids Encoding Sam
; FILE REFERENCE: 15966-67CIPCON1
; CURRENT APPLICATION NUMBER: US/10/150,811
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 09/970,607
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/182,733
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/182,724
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 60/183,896

```

```

; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: 60/184,497
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/224,157
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/184,482
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,744
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/197,083
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/233,405
; PRIOR FILING DATE: 2000-09-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-150-811-70

Query Match      57.7%; Score 8974.5; DB 12; Length 3034;
Beet local Similarity 57.3%; Pred. No. 0;
Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;

1  MRSPTAVPLPPTPPPLLLLLLLLLPPPLGDDVGPGRSLGSRGSG-----S 48
Db      124  LRRSARGAELRSP-----AVRSVPGLDAL--CFPAAGGAAALVSLBAITWPPA 172
Qy      49  GACAPM-----GWLQ--PSSASNLMTYTSRCADAGTELTGHLVPHHDGLRVMCPSEA 99
Db      173  CSCFPVAGTCGRRPICLRPGSGAELALVCLGAQA-----VWV----- 213
Qy      100  HPLPAPBEGCPWSCRLIGIGHLSPQKLTLPENHCLAPRLRCOSCKLAQAPGLRAG 159
Db      214  -----ELVIGATSGTPBSPSV--SPSL-----LWLSQP--RAG 243
Qy      160  --BRSPEESIGGRKRVNTAPQGPSYQATVPENPAGTPVASIALIPDBEAGRL 217
Db      244  VRRS-----RGTGSSSTSPPLPSYQVSPENPAGTAVTELAHPDEDEDAKRLS 296
Qy      218  TYMDALFDSRNOFFSLDPTGAVTTAELDRKSTHYVRVTTAODHGMPPRSALATLTI 277
Db      297  YQMALFDERNSNGFLIDATGAVTTARSIDRETQTHVAKVASVDGSPRRSALATLTI 356
Qy      278  LVYDNDHVDVFFQOEYKESLRNLEVGIVLTVRATGDAPEANAILYRLBSSGSSPS 337
Db      357  TVSDPTNDHSFVFQOSEYRERIRNLEVGIVLTVRATGDAPEANAILYRLBAGG--- 413
Qy      338  EVPEIDPRSGVIRTRGVDEEVEBSVOLTYEASDQGRDPGRSTTAVFLSVEDNDNAP 397
Db      414  VFEIDARSGVAVTRAVVDEEBAEYQLVLEANDQGNPPLSASATVHIVVEDENNYR 472
Qy      398  QFSEKRYVQVRDVTGPAVLRTASDRDKGSAAVHVHIMSNGARQGYLDAQTGALD 457
Db      473  QFSEKRYVQVRDVTGPAVLRTASDRDKGSAAVHVHIMSNGARQGYLDAQTGALD 532
Qy      458  VSPFLDYETTKETTLRVADGGRPPLSNVSLVTVQVLDINDNAPFVSTPFOATVLES 517
Db      533  VINPLDPEALREYTLAKQDGRPPLINSGLVSVQVLDINDNAPFVSSPQAAVLN 592
Qy      518  VPLGVTLVHQAIDADAGNARLEYRLAGH-----DPEPTINNGTG 560
Db      593  VPLGVTLVHQAIDADAGNARLEYRLAGH-----DPEPTINNGTG 652
Qy      561  WISVAELDEEVDYFSFGVEARDHGTPALTAASVTVTLDVNDNPPFTQPEYTVRLN 620
Db      653  WITVCALDREBEVHSFQVEAVDHSPPAMSSASVITLVDVNDMPMTQVYELRLN 712
Qy      621  EDAAGTVSVTVAVVDRDAHSVTTQITSGNTNRRSITSQSGGGLVSLALPDYKLERQ 680
Db      713  EDAAGTVSVTVAVVDRDAHSVTTQITSGNTNRRSITSQSGGGLVSLALPDYKLERQ 772

```



Db 2904 SYELHROAGNHCGRPSDPSGVLA-----FVAVLSQPOBQKGLKKNVTPPLP 2958  
Qy 2849 EK--SSLRLPEOCTGSSRG-SASBGRGPP-----RPPROSLQOLGVMP 2898  
Db 2959 EOPLRKREKLADQSFSTSRSLSGDGVHATDVTITKTRREREGREHLNGV--- 3015  
Qy 2899 AMSIXAGTVDEDSGSE 2915  
Db 3016 AMNVRTGSAQNGSDSE 3032  
RESULT 13  
US-09-737-149-2  
; Sequence 2, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shinkels, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Spyrek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT APPLICATION NUMBER: US/09/737,149  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,544  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 2000-01-04  
; PRIOR APPLICATION NUMBER: 60/174,962  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 3014  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-737-149-2  
Query Match 56.3%; Score 8754; DB 9; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  
Qy 12 TPRPRLILLLILPRLIGDVGPRSLGSRGSSGACAMKLCPSASNLMLYTSR 71  
Db 57 TPRARELL-----DVGGRDGLAGR-REVSGAGRL-----PLQVRLV----- 93  
Qy 72 CADAGTEYLGHVPHNDLGRWCSESEAHIRPRABEGCWSCLLIGGHLSPQKLT 131  
Db 94 ARSATLALSRL-----RAKTLR-----GCGARALCGTARIC--GALCF 133  
Qy 132 P-----EENPCLKA-----PRLRCOS-----CKLAQAG- 155  
Db 134 PTVGGCAAAQHGALAAFTLLPACRCPPRRPRRCPPGICLPFGSGVRLRLCALARAAGA 193  
Qy 156 LKAG-----ERPEESLGGRRKRVNTLP-----QFPPSYQAT 189  
Db 194 VAVGALAEATAGTPSASPSPPRLPRLPEARAPARARARGTSGRLKPMENYQVA 253  
Qy 190 VENOAGTPVASTLAIDPDEGAGRLLEYTMDALPDSRNOFPSSLDPTVATTAEBDR 249  
Db 254 LPEENPAGTLILQLAHAYTIBGEBEAVSYTMEGLPDERSGYFRIDSATGAVSTDSVDR 313  
Qy 250 ETKSTHVPFVTAQDHGMPRRSALATLTTLVTDTHDVPFEOEYKESLRENTLEVGVEVL 309

Db 314 ETKTHTLRAVAADVSTPPSAGTATYITVLVYDTHDHPVFEOQSTRRERENTLEVGVEVL 373  
Qy 310 TVRATDGAPEANILYRLLEGSGSPSEVEFIDPRSGVIRTRGPVDRBEVSYQLTVEA 369  
Db 374 TIRASDRDSPINATLRYRLVGAW-----DVQLNESSGVVSTRAVLREBAEAYQLVEA 429  
Qy 370 SDOCRDGPSTTAAVFLSVEDDNDNAPQPSERKYVYQVREDYTPGARVYLAVTASDDDKG 429  
Db 430 NDGRNPGPLSAIVATVYIIEVEDENDNYPQPSBQNYVQVPEPDLNATVLEQVATHDQ 489  
Qy 430 SNAVHTSINGNAGQFYLDACQALDVVSPLDYETTKETKYLAVRAODGGRPLPSNVSG 489  
Db 490 QMAIHYSIISGNVAGQFYLSLSGLIDVYNPDLFEDVQKTSLSIKQDGRPLPILINS 549  
Qy 490 LVTYQVLDINDNADIPYSTPPQATVLESVPLGYVTHVQALDADGNARLEYLA---- 545  
Db 550 VVSQVQLDVNDNEPIFVSSPFOATVLENVPLGYPVVHIOAVDADSGENARLHYRLVDTAS 609  
Qy 546 ---GVGH-----DPEPTINNGWTSVARELREEDVPFSGVEARDHCTPALTA 592  
Db 610 TFLGGSGAGPPNAPPTDPFPQIHNSSGWTVCALDREBEVHYSFGEAVDHOSPPMS 669  
Qy 593 SASVTVTLVDVNDNPTFTQPEYTVRLNEDAAVSTSVVTSADVADHASVITYQITSGT 652  
Db 670 STVSITVLDVNDNDPFTQPTVELRLNEDAAVSSVLTICARDNDANSVITYQITGNT 729  
Qy 653 RNRFTSISQSGGIVSIALPLDYLERQYVLAVTASGTRQDTAQIYVNTDANTHRPVF 712  
Db 730 RNRPLSSORGGIITLALPLDYERQYVLAVTASGTSHTAHVILVNTDANTHRPVF 789  
Qy 713 QSSHVTYVNDNRAGTIVLISATDDEGENARITTFMBSIPQFPIIDVGTAVTQAE 772  
Db 790 QSSHVTYVNDNRAGTIVLISATDDEGENARITTFMBSIPQFPIIDVGTAVTQAE 849  
Qy 773 LDYEDVSYTLATATARDNGIPQKSDITYLRLVNDVNDNAPOFLRDSYQSGVEYDVPFT 832  
Db 850 LDYENQVATYLTIAQNGIPQKSDITTLRLILDANDNAPQFLMDPYQSGIFEDADPST 909  
Qy 833 SVLQISATDSDSGANGVFTTQGGDGGDDPFYESTSGIVRTLRILDRENAQVYLRAY 892  
Db 910 SILQVSATDRDSGNGGLTYTFOGGDDGDPFYLRPSGVIRTORRLDRENAVAYYLMAL 969  
Qy 893 AYDKGMP-PARTPEVTVTVLDVNDNPVFEODEFDFVENSISGLAVARVATDPPDG 951  
Db 970 AYDKGSPPLSASVEIQVTLIDINDNAPMEKDELELFEENNVGSYVAKIRANDPDEG 1029  
Qy 952 TNAQIMQIYEGNIPVFOULDIPSGELTALVDLYEDRPEYVLIQATSAPLVSARATVH 1011  
Db 1030 PNAQIMQIYEGNIRHFFQDLNLNGDLRAVAVELDFVRAREVYLVQATSAPLVSARATVH 1089  
Qy 1012 RLDRNDNPYLGAFELLFNNTYNNRSSPFGAIGAVPAHPDISLTSYSPFGNELS 1071  
Db 1090 LLVQNDNPPVLPFOILFNNTYNNKNSPPTGYIGCPADHPVDSILNTYVQSGNELR 1149  
Qy 1072 LVTLNASTGELKLSRALDNNRPLEAIVSVSDSVSTVQACALRYITIDEMLTHSITL 1131  
Db 1150 LILNDPATGELQLSRDLDNNRPLEALMEVSDSHVTAFTIKRVITITIDMLTNSITV 1209  
Qy 1132 RLBDMSPERFLSPLGLFIQAVATLTATPDHVVVFNVOBTDAPGGHILNVLSVQOP 1191  
Db 1210 RLBNMSQEKFLSPLALFVGVAAVLTSTKDDVVFVNVQDTDV--SSNITLVTSALLPG 1268  
Qy 1192 GPGGPPPLPSBEDIQEYLINRSLTLTASQRYLPNDNICLAREPCENYMCVULAFDS 1251  
Db 1269 GVRG--QFPSEEDIQEOLYINRSLTLTISTQRYLPNDNICLAREPCENYMCVULAFDS 1336  
Qy 1252 SAPFIASSVYLFRIHVPVGLRCPCPGFTGADYCEYVDICYSRPPGPHGCRREGGYT 1311  
Db 1327 SAPFLSSTVYLFRIHINGLRCPCPGFTGADYCEYVDICYSRPPGPHGCRREGGYT 1386  
Qy 1312 CLCRDGYTGEHCEVASBSGCTPGVCNNGGTGVNLLVGFGRKDCPSGDPEKPYCQVTRS 1371



1387 CECFEDFTGHCEDVARSGRGANGCKNGGTGVNLLIGFPHCVCPPEGEYRPYCEVTTSS 1446  
QY 1372 FPAHSPTFRGRLRORFHFPTLALSFATKERDGLLYNGRFNEKHDPALEVIQOVLTTSS 1431  
DB 1447 FPOQSVTRGRLRORFHFPTLALSFATKERDGLLYNGRFNEKHDPALEVIQOVLTTSS 1506  
QY 1432 AGESTTVSBPVGVSDQGMHTVOLKYXNKPPLAQGTGLPOGSEKQVAVVTVGDCDTGV 1491  
DB 1507 AGESTTVSBPVGVSDQGMHTVOLKYXNKPPLAQGTGLPOGSEKQVAVVTVGDCDTGV 1566  
QY 1492 ALARSGVLNYSGLAAGTQGSKSLDLTGPPLLGVPDLPESEFVYRMRQFVCNENLOY 1551  
DB 1567 AAFEGDIGNYSGLAAGTQGSKSLDLTGPPLLGVPDLPESEFVYRMRQFVCNENLOY 1626  
QY 1552 DSHHIMADFIANGTVPCCPAKGVCDSTNGGNGTGVNQMADAFGCEGLGFGKSCAO 1611  
DB 1627 DGNQVMAGTIANNGTREGCMARNTCDGRCONGTGVNRMNTYCECEPLRGKXNQEQ 1686  
QY 1612 EMANPQHLGSSLVANHGSLPISQPMYLSLMFRTRQADGVLLQAITRGRSTITLQLRBG 1671  
DB 1687 AMHPQLFSGESVYMSDNLNIIISVPMYGLMFRTRKEDSVLMEATSGGPTSRLLQILNN 1746  
QY 1672 HYNLSTEGTGLQASSRLBERGRANDGMHQAOL--GASGPRGAL-ISFPGQORAE 1727  
DB 1747 YLQFEVSHGSDVESYMLSGRLVTDGMHLLIETLNKXKEDSEMKHLYMTTLDGMDQNK 1806  
QY 1728 GNLGPRHLGLHLNITVGGIRGPAAGVARGFCLOGVRSYDTPREGNLSLDSHGSINY 1787  
DB 1807 ADIGMLPGLTVASVYVVGASBDRKSVYRGFRGCMGVNKGTFPTVATLMMNNALKVRY 1866  
QY 1788 BQGSCLPDCSNPCPANSYCSNDMDSYSCSDPGYGDNCVTNCDLNPCEHOSVCTRKP 1847  
DB 1867 KQCDVDVDDPCTSSPCFPNSRCHDAMEDYSVCCKGTLGNCVDAQHLNCPENNAGACVRSR 1926  
QY 1848 SAHGTTCBPBPYNYLRYCETRIDOCBPGRMGHPTCPGPNCEVSKGFPDCKTSGEGR 1907  
DB 1927 GSHQGVYCEGSHYBPCENKLDLPCPRGMGNPCVGPCHCAVSKGFPDCKTSGEGR 1986  
QY 1908 CKNHRYRPPGSPYCLLCDYPTGSLSRVCDPEPDQCPCKPGVIGRCCDRCDNPPAEVTN 1967  
DB 1987 CKENYKTLAODTCLPCDPCPHSHSRITCMATGQCAKRGVIGRCCNCDNPPAEVTN 2046  
QY 1968 GCEVNTDSCPRALBAGIWMPTRPGLPAAAPCEKSGFNAVRHCDHRGMLPNNFNCTG 2027  
DB 2047 GCEVNTDSCPRALBAGIWMPTRPGLPAAAPCEKSGFNAVRHCDHRGMLPNNFNCTG 2106  
QY 2028 ITSEELKGFARLQNRSGSLDSRSQLALLANATQHTGSGVYKAYOATLTLAH 2087  
DB 2107 ISFVDLRANNEKLSRNETOVDGARALQVRLASATQHTGTLFGNDVTRAYOLLGHVLR 2166  
QY 2088 ESTQRFGLSATQDVHFTENILRVGSLALDTANKRHWEILQTEGGTAMLLQHYEAYASA 2147  
DB 2167 ESTQRFGLSATQDVHFTENILRVGSLALDTANKRHWEILQTEGGTAMLLQHYEAYASA 2226  
QY 2148 LAONMHTVYSPPTVTVNIVISVRLDKNPAKALPRYALRGSEOPDLFTTVILPES 2207  
DB 2227 VAENVRRTYLRPVIVYANMILAVDIFDKENFTGARVPRDITHEEPRELESSEVFPAD 2286  
QY 2208 VFR-----ETPVVRPPG-----RGEAQBPEELARQRRRRPELSQGEAVASVITRT 2254  
DB 2287 FFRPPEKEGRLRPPARRTTTPOTTRPGPOTERBAPISRRRRRPDDAGQFAVALVITRT 2346  
QY 2255 IAGLLPHNYDDPKRSLRVPKRPINTPEVVISVHDEBELPRALDVKVTVQFRLLETER 2314  
DB 2347 LGQLLPRYDDPKRSLRVPKRPINTPEVVISVHDEBELPRALDVKVTVQFRLLETER 2406  
QY 2315 TXR1CVFMMHSILVSGTGWSSARGCEVFPNESHVSQCNHMTSPAVLNDVSRRENGELL 2374  
DB 2407 TXR1CVFMMHSILVSGTGWSSARGCEVFPNESHVSQCNHMTSPAVLNDVSRRENGELL 2466  
QY 2375 PLKTLTVAGLTAALLTLFPFLLTLRLIRSNQHGIRMLTAALGLAOLVPLGIGNOD 2434  
DB 2467 PLKTLTVAGLTAALLTLFPFLLTLRLIRSNQHGIRMLTAALGLAOLVPLGIGNOD 2526

QY 2435 LPFACVTAIILHPLYLCTFSWALLLEALHYRALTEVDVNTGPMRFYTMGMGVPAFT 2494  
DB 2527 NFPLTVAILLHPLYLCTFSWALLLEALHYRALTEVDVNTGPMRFYTMGMGVPAFT 2586  
QY 2495 GLAVGLDDEGYNDPFCWLSYDYLIMSPAGVAVASVVELYLLAARASCAORQFBE 2554  
DB 2587 GLAVGLDDEGYNDPFCWLSYDYLIMSPAGVAVASVVELYLLAARASCAORQFBE 2646  
QY 2555 KKGVSGLQSPAVULLLSATWLLALSVNSDTLLFHLFATCNLCIOGPFITLSVYLSK 2614  
DB 2647 KKGVSGLQSPAVULLLSATWLLALSVNSDTLLFHLFATCNLCIOGPFITLSVYLSK 2706  
QY 2615 EYRKALK-LACSRKSPDPALTTKST-LTSSYNCPSPADG--RYOPFGDAGSLHSTS 2670  
DB 2707 EYRKALKVLGGRKHLHEDSATTRATLTLRLSNCTMTGDEGDMRTLGSTABLDSTIV 2766  
QY 2671 RS-----GKSQPSYIPFLLRBSBALNPGQPPGLGDPGLFLEGQDQ 2713  
DB 2767 RDEGIQKLGVSGLVRSRSHGEPDASLMPRSCD-----PPG----- 2802  
QY 2714 HDPDSDSDSLDEPDQSGYASTSSDSSEBEEBEEBAAPRGQWDSILGPABRLP 2773  
DB 2803 --HDSDSSELSL--DEQSSSYASSHSSDSEDDGVABERK-----WDPARG-----A 2845  
QY 2774 LHSTPK-DGPGPGRAPWGPDPGTAKESGNGAP-----BEERLRENGDAL 2818  
DB 2846 VASTPKGAVANHVAVGMPDQSLASDSEBPGKRLKVTYKVSVELRREBQSHRGYR 2905  
QY 2819 SRESGLPPLGSSAOP--HGKILKKCL--PTISKS--SLRLPLEQCTGSSRGS 2869  
DB 2906 PQESGGAARLASQPPRGKILNKTYPPLTLTQTLKGRREKLADCEQSPTSR 2965  
QY 2870 ASESGRGP---PPRPPROSLQBLNCGWPMIANSIAGTVDESSGE 2915  
DB 2966 TSLSGGPDCAITVKSFPERGRDLNGLV---ANAVRTGSAQADGSDSE 3012

## RESULT 14

US-10-241-220-107  
Sequence 107, Application US/10241220  
Publication No. US20030148408A1  
GENERAL INFORMATION:  
APPLICANT: Frantz, Gretchen  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Phillips, Heidi  
APPLICANT: Polakis, Paul  
APPLICANT: Spencer, Susan  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wu, Thomas  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
TITILE OF INVENTION: TREATMENT OF TUMOR  
FILE REFERENCE: P5010R1-US  
CURRENT APPLICATION NUMBER: US/10/241,220  
CURRENT FILING DATE: 2002-12-13  
NUMBER OF SEQ ID NOS: 120  
SEQ ID NO 107  
LENGTH: 3014  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-241-220-107

Query Match 56.3%; Score 8754; DB 12; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;

QY 12 TPEPPLLLLLLLLPPLLGQVGPGRSLGSGRSGSACADMGWLCPBSSANLMLYTSR 71  
DB 57 TPRAPRELL-----DVGRDGLRAGR--RVSAGAGRP-----PLQVRLV----- 93  
QY 72 CNDAGTELHGLVPHHNDGLRVWCPRSESAHILPRLPABCPMSCRLLGTGCHLSPGKLT 131

Db 94 ARSAPATLNRRL-----RARTILP-----GCGARARLCGTGARLC--GALCF 133  
 Qy 132 P-----BEHPLCKA-----PRLRCQS-----CRLAQAPG- 155  
 Db 134 PVBGGCAAAQHSLAAPTLLPACRCPPRRPRRCGRPICLPBGSGVRLMLCLRLRBAQA 193  
 Qy 156 LRAG-----ERSDESLGRRKRNVTAP---QOPPSYQAT 189  
 Db 194 VVRGALAEATAGTPSASPSPPPLPPNLPBARACGARRARRGTSRGSLKCFMPMYQVA 253  
 Qy 190 VBNOPAGTFVASIRAIIDPDEGAGRELEYTMDALPDRSNOQPSLDPVYGAUTLAELLR 249  
 Db 254 LRENEBAGTLIIQJAHYHITIEGSEKVSYSMEGLPERSRGYFRIDSAVGAUSTDVLDR 313  
 Qy 250 ETKSTHVFVTAQDHQMPRRSALATLLIIVTDNTNDPVPBQOEYKESIRENLEVYEL 309  
 Db 314 EFKETHLVKAVDVSTPPRSATITIVLYKOTNDSPVFEGSEYERERENLEVYEL 373  
 Qy 310 TYRATDGDAPRNANILIRLBSGGSPSVFELIDPSGVIRTRGVDRREVSQUTVRA 369  
 Db 374 TIRASRDRBPINAMIRYVLGGM---DVFOINBSGVSSTAVALDREBALEYQLVBA 429  
 Qy 370 SPOGRDPRGSTTAAVFLSVEDNDNAPOPSKRYVVOYREBDVTPCAPVLTATASDRDG 429  
 Db 430 NDOGRNPGLSATATVYIEVEDENDNYPOPSQONVVOYVEDGANTAVLRVQATDRDG 489  
 Qy 430 SNAVHTSIMGNARQFYLDAGTALDVSPLDYETTEKYLTVRAODGGRPLSNVSG 489  
 Db 490 QNAALHYSITLSCGNVAGQFYLHSLSGILDVYNPLDPEDVQKTSIKAQOGGRPLINSQ 549  
 Qy 490 LVTYQVLDNDNAPFVSTFPOATVLESYPLGLVLAHVQALDADGNARLEYRLA--- 545  
 Db 550 VVSVOVLVDNDNEPIFVSSBPQATVLENVPLGYPVHIQAVDADSENNARLEYRLVDTAS 609  
 Qy 546 ---GVGH-----DEPTINNGTGMISVAABEDREXVDFYSGVAREBDHGTALTA 592  
 Db 610 TILGGSGAGPKPAPRPPDPFOIHNSSGHTVCAELDRREXVDFYSGVAREBDHGTALTA 669  
 Qy 593 SASVSVTVLDVNDNPTPTQPEYVALNEDAAVGTSVTVSAVDRDASHVITYOITSGNT 652  
 Db 670 STSVSITVLDVNDNPPVFTQPTVELRLNEDAAVGSVTLQARDNANSVITYOITSGNT 729  
 Qy 653 RRRPBITSGGGGLVSLAPLDYKLEKQVLAVTASDGTRODTAQIVANNVTANTHRPVF 712  
 Db 730 RRRFALSSORGGGLTTLALPLDYKQEQYVLAVTASDGTSHAVLINVDTANTHRPVF 789  
 Qy 713 QSSHATVANNEDRPAGTIVLISATDEBGENARITYPMEDSIPOPRIDADGAVTQAE 772  
 Db 790 QSSHITVANNEDRPAGTIVLISATDEBGENARITYPMEDSIPOPRIDADGAVTQAE 849  
 Qy 773 LDYEDQVSTTLAITARANGIPOKSDITYLEILVNDVNDNAPOLFRLDSYQGSVYEDVPPPT 832  
 Db 850 LDYENQVAVTTLTIMADNGIPOKSDITYLEILVNDVNDNAPOLFRLDSYQGSVYEDVPPPT 909  
 Qy 833 SVLQISATDRBSGLNGRVYTFQGGDDGDDSTVESTGIVTLRLDRBNVQVYLAAY 892  
 Db 910 SILQVSAIDRBSGPNRGLLYTFQGGDDGDDGYIAPTSGVIRTORRLRENAVAVNLVAL 969  
 Qy 893 AVDKGMP-PARTPMEVTVTVLDVNDNPPVFEODEBDFEYBENSPIGLAVARTATDPDG 951  
 Db 970 AVDRSPFPLASVETIQTIIIDINDNAPFEKDELEFVEENPVGSVAKIRANDPDG 1029  
 Qy 952 TNAQIMYOIVEGNIPEVFOLDIFSGELTALVDLYEDRPEYLVIOATSAPLVSATVHV 1011  
 Db 1030 PNAQIMYOIVEGDMHFPQJDLINDLRLAMVELDEVERREYLVIOATSAPLVSATVHI 1089  
 Qy 1012 RLDDNDNPPVYAGNEILLNNVYTRSSSPFGCALGRVPAHDPDISDGLTYSFREGNLS 1071  
 Db 1090 LLDVNDNPPVLPDQOILFNNAVTKNSFPVGVGCIPAHDPDVSISLNTYFVGNBIR 1149  
 Qy 1072 LVTLASTGELTSLALDNNRPLEAKIMSVLSDGHSVTAOCALVTTITBMLTHSITL 1131  
 Db 1150 LLLDPAIGEQLSDLDNNRPLEALMRYSVSDGHSHTVAFCTIAVTTITTDMLTNSITV 1209

Qy 1132 RLEDSPERFLSLGLFLIOAAVATLATPPDHVVFNVOJRDADAGHILANLSVGOPE 1191  
 Db 1210 RLENNQKFLSLPLALFVBGAVALSTTKDDVFPFNQNDTV--SNILANTFSLALRG 1268  
 Qy 1192 GFGGPPPLPSDLOERLYNLSLITAI SAORVLPDDNI CLRPCEYMCVSVLRPDS 1251  
 Db 1269 GVRG--QFPSEBDLBOQYLYNRITLLTISTQVLPFDNI CLRPCEYMCVSVLRPDS 1326  
 Qy 1252 SAPFLASSVLPRIHPVVGARCRCPPTGDCETVLDLCSRCPGPHRGSRREGYT 1311  
 Db 1327 SAPFLSSVTVLPRIHPVVGARCRCPPTGDCETVLDLCSRCPGPHRGSRREGYT 1386  
 Qy 1312 CLCRDGYGHCCEVARSARSGCTPVCNKGCTVNLVAGFKDCDSSGPEKPYCOVTRS 1371  
 Db 1387 CECFEDFPGHCEVDARSGRCNKGCTVNLVAGFKDCDSSGPEKPYCOVTRS 1446  
 Qy 1447 FPPQGFVTFRGIRORFHTLITLTFATQBRNGLLYNGRFNEKHDFIALEIVDEQVLTFS 1506  
 Db 1432 AGESTTVSPVPGVSDGOMHTVQAKYNNKPLGQGTLPQGSROKXAVVTVDCDPTV 1491  
 Qy 1507 AGEITTVAPKVPVSGVSDGRHNSVQVYNNKPNIGHLGHPGSGEKXAVVTVDDCDPTM 1566  
 Db 1492 ALRFGSVLGNTSCAAQGTQGSKSLDLTGPLLGGVPLDPESFVVRORQFVGCRRNTQV 1551  
 Qy 1567 AVRFKQDIGNVSCAAQGTQGSKSLDLTGPLLGGVPLDPESFVVRORQFVGCRRNTQV 1626  
 Db 1552 DSRHIDMADFIANNGTVPBCPAKKNVCDNNTCANGGTVCNQMDFSCBCLPFGGKCAQ 1611  
 Qy 1627 DGKNVDMAGFIANNGTVPBCPAKKNVCDNNTCANGGTVCNQMDFSCBCLPFGGKCEQ 1686  
 Db 1612 EMANPOHFLGSSIVAMHGLSLPISOPWTLSPMPTROADGVLQAITRGSTITQLREG 1671  
 Qy 1687 AMPHOLEFSGSVVSSDNLITISVPWTLGMPFRKEDSVLMKETSGLPYSFRQIILNN 1746  
 Db 1672 HVMLSVEGTGLQASSLRLPERGRANDGMHHAQLAL--GASGGGHAL--LSFDYGOQRAE 1727  
 Qy 1747 YLQFVSHGPDVDSVMSLGRVTDGEMHLLILKRVKEDSBMKHLVMTLTDVGMQNK 1806  
 Db 1728 GNLGPRHLGHLNSITVGGITGPAGGVARGRGCLQGYRVSDTDEGVNSLDPHGESINV 1787  
 Qy 1807 ADIGGMLPGLTVRSVVVGASBDKVSVARGRGCMQGRMGCTPTNVAITLMNNALAKRV 1866  
 Db 1768 EOGSLPDPCCNSNCPANVSYSNDMDYSQSCDPTGYGDNCTVNCDDLPCHEQSVCTKRP 1847  
 Qy 1867 KQCCDVDDPCTSSPCPPNSRCHDMEDISCVDKGYGLINCVDACHLNPCEMGA CVASP 1926  
 Db 1848 SAPHYTCECPNPIAGPYCETRIDQPCPRGMWGHPTGCPNCNDVSKGDPDCNKTSGECH 1907  
 Qy 1927 GSPQGYCEGSPSHYGPYCEKMKDLPCPRGMWGNPVGPGCHCANVSKGDPDCNKTNGCQ 1966  
 Db 1908 CKENHYRPGSPPTCLDCTYFSGLSRYCDPEDQCPCKPQVYTRQCDRCNPNPASTTN 1967  
 Qy 1987 CKENYKXLAODTCLPCDCPFGHSHRTCDMATQCAKCPGVIRQCCNCPNPAEVYTL 2046  
 Db 1968 GCRYNVDSCPRAIAGIWPPIRFGLPAAACCPGSGSGTAVRHCDEHGMPLPMLFNCTS 2027  
 Qy 2047 GCERYNVDSCPRAIAGIWPPIRFGLPAAACCPGSGSGTAVRHCDEHGMPLPMLFNCTS 2106  
 Db 2028 ITFSELGFAERLQRNESGLDGSQQLALLRNATQHTAGYFGSDYKVAVOLATRLLAH 2087  
 Qy 2107 ISFVLDLAMEKISLRNETQYDGAVALQVVALRSATQHTGTLFNDVRAVAYOLLGHTLOH 2166  
 Db 2088 ESTORGGLASTOQVHTENMLRVGSLALDTPANGHHELIQOTGCGTAMLOHBAASA 2147  
 Qy 2167 ESWQGFPLAATQADHAEVHSHSGSALLAPATYAAEQIORSSEGTQOLRLRLEGYFSN 2226  
 Db 2148 LAQNNRHTVLSPPFITVPNIVISVVRLDKGNFPAKAPRYEALAGEOPDLETTVLPES 2207  
 Qy 2227 VARVVRITVLPFVIVANNMILANDIFDKFNFGARVPRPTIHBEPRBLSSVSFPAD 2286



Qy	1012	RLDRLNDPRLVGNFELLPNNTYVTRSSSPFGCAGVPAHPDPIISLJYTSFERGENLS	10711
Dp	1090	LLVQNDNBPVLPDQOILFNNVTVMKMSFPFGVIGCIPAHDPBVSISLNTFYQGENLR	11434
Qy	1072	LVLTANASGEJLKSALDNNRPLBAIMSVLSDGHSVTAQCALVITIITBMLTHSITL	11133
Dp	1150	LLIIDLPAIGEQLSDLDNNRPLELMEVSVSDGHSHTAFCTLLVITIITDMLNHSITV	12093
Qy	1132	RLDMSPERPLISPLIGLFIQAAVATLATPDPHVVFVFNVQORDTABGHIINVSLSVGOP	11191
Dp	1210	RLNNMSQEKFLSPILLALFVEGVAAVLSTTKDVFVFNVQNDTIV--SNSILNVTFSALLPG	12668
Qy	1192	GGGGPRLPLBBDLOERLYINRSLLTALSAGVLPFPDNICLARECENYMCVSYLARDS	12515
Dp	1269	GVRG--QFPSSBDLOEQIYLNKTLITLSTQVLPFDNICLARECENYMCVSYLARDS	13262
Qy	1252	SAPFIASSSVFRRPIHPVGLRCRCPPTGDCYCEVEVLDLCSRPCGPHRCRSHEGYT	13111
Dp	1327	SAPFIASSVTVLFRRPIHPIINGLRCRCPPTGDCYCEVEVLDLCSRPCGANGCRSHGYT	13669
Qy	1312	CLCRDGYTGEHCERSASRGCTPGVCNKGCTCVNLLVGGFKDCDPSGDPFKPYCOVTRS	13711
Dp	1387	CECEDPFGHECEVDARSGRCANCGCKNGCTCVNLLIGFHCVCPCGHEYERPYCEVTRS	14466
Qy	1372	FPASFIIFRGRLROHFHTLALSPTKRDGLLYNNGFNRKHPDVALEVIQEOVLTRS	14311
Dp	1447	FPPOGFIYFRRLRQRHFHTLISITPAQORNGLLLYNNGFNRKHPDVALEVIQEOVLTRS	15066
Qy	1432	AGESTTVSPVPGVSDGQWHTVQJKNRPLLGQTSLPGSPBQKVAVTVDGCCTGV	14911
Dp	1507	AGESTTVTAPKPVSGVSDGRMHSVQVYKXNKPNIIGHLGLPHGSPGKMAVTVTDCCDTM	15666
Qy	1492	ALRFGSVLGANSCAAQGTQGSKSLDTGPLLGGVVDLHPSPPVNRKROVGCNRLQY	15511
Dp	1567	AVRFGKDJGANSCAAQGTQTSKSLDTGPLLGGVNLBDFPVNHRQVPGVCNRLSV	16262
Qy	1552	DSRHIDNADFIANNGTVPGCAPKKNVCDSNTCHNGTICVNOMDAPSCCPLAFGSGKCAQ	16111
Dp	1627	DGKVDNADGFIANNGTREGCAARNFCDGRCCQNGCTCVNMMNYLGCPLRFGKNCBQ	16866
Qy	1612	EMARPOHLGSSVLAHNLISLPISQPWYLSLMPFRKADGVTLQAITRGSRTIQLRBC	16711
Dp	1687	AMHPOLPSSGSSVWSMDLNTIISVPMYLGIMFRKRKDSVLMKRTSGSPITSFRIQIINN	17466
Qy	1672	HVMLSVEBGTGQAASILRLPERANDGDVHHAQLAL---GASGGPFAH--LSFVGGQGRAB	17271
Dp	1747	YLQREVSHSGPDSVESVMLSGLRYVDGBHHLLILKXNKEDSEMKHLVMTYMLDGMQNK	18066
Qy	1728	GNLCPRHLGHLNSITVGGIDPGAPAGVARGGCLQGVRSVDTBEGVNSLDSHGESINV	17871
Dp	1807	ADIGMPLGLTVRSVVVGASBBDKXSVARGFRGCMQGRMGCTPTNVAITLMMNNAKLKRV	18666
Qy	1788	EGGSLPDPCCSNPNCANSYCSNMDSDYSGCDPGYVDNCVMYCDLNPCEHOSYCTRKP	18447
Dp	1867	KDCGCDVDDPCTSSPCCPFRNSCHDAMEDYSCYCDKGYLGLNLCYDACHLNPCEMGAQVRSR	19262
Qy	1848	SAPHGYTCECPNYLGPYCETRIDQPCBRGWMGHPTGCPNCNDVSKGFPDPCNTKISGECH	19071
Dp	1927	GPSGQYVCECGPSHSGPYCENKLDTLPCPRGWMGNVGPCFCHCAVSKGFPDPCNTKNGCQ	19666
Qy	1908	CKENHYRPGSPCTCLCDYCPYGLSVYCDDBDQCPCKBPVITRQCDRCNPPRAYVTIN	19667
Dp	1987	CKENYIKLIADPTCLPCDCFPHGSISRCDMAQCAKCPVIGIQCCRCNCPRAVETTL	20466
Qy	1968	GCEVNYDSCPRAIENGIWPFRTREGLPAAACCPKOSFCTAVRHCDENHGMPLPNIJNCTS	20271
Dp	2047	GCEVYINGCPRAPEGIWVQTKFQPAVAPCPKGSVGNAVRHSGEKGMPLPELFNCTT	21066
Qy	2028	ITTBELKGFARLQNESGJLDSGSQLALLNRYAOTHTAGYRPSDYVAVYQATRLIAH	20871
Dp	2107	ISFVDLRAMNKLSRNETQVYDGAARLQVLRALRSATQHTGTLFGADVATAQOLLGHTVQH	21666

QY	2088	ESTORGSLSTOQVHTENTLIRGYSLLDPTANKRHMLLOQTEGSGAMLLQHTBAVASA	2147
Db	2167	ESMOGEDLMAITQADFHEDVITHSGSLLBPATBAEBOJORSBGGINOLRLIEGIFSN	2226
QY	2148	LAONMRHTYLSPTTIVTENIVISVRLDKGNFAGAKLPRYEALNGEOPDLETTVIIIPES	2207
Db	2227	VARNVRETYLRPFVIYVANNMLAVDIEDKFNFTSARVBERPDTIEEPRELESSVSPAD	2286
QY	2208	VFR-----ETPVVRPAG-----PGBAOPEBELARQORRHLESGEYVAVITYRT	2254
Db	2287	FFRPPEBEGGLPAPRRRTTPQTRGCGTERAPISRRRRHPDDGQFVAALVITYRT	2346
QY	2255	LAGLLPHNYDPDKKSLRVPKRPITITNVSISVSHDEEELLPRADLKVTVOFRLLTEER	2314
Db	2347	LGOLLPRYOPDRRSLLPHRPINTPMWSTLYVSEBAPLPRLPRVIVVAFALLVBER	2406
QY	2315	TKPICVPMNHSILVSGTGVMSARGCEVVFNRSHVSCQCNHMTSPFVLMVDSVRNENGIL	2374
Db	2407	TKPVCVPMNHSILVSGTGVMSARGCELLSNRRTIVAOCCHSTASPAVIMDISRENGEVL	2466
QY	2375	PLKLTVVALGVTLTAALLTPFFLTTLRIIRSNQHGIRNMLTALGLVLAOLVFLGINQAD	2434
Db	2467	PLKIVTTAAVNSLSIALMLVAFVLSLVRMLRSNLHSIHKILAVALLFSOLVFTVGINQTE	2526
QY	2435	LFPACTVIAILLHFYLCTFSWALLBMLHYRALTEVRDVNTGEMRPFYMLGWMVPAFIT	2494
Db	2527	NPFICTVVALILHRIYVSTPAMTLVBSLAIYRMLTEVRNIDTGMRFYVYVGMGIPALVT	2586
QY	2495	GLANGIDPBYGNDPCWLSIYOTLWSPGPAFAVMSVFLTYIAARASCAARQGFGE	2554
Db	2587	GLAAGIDPQGGNDPCWLSLOTLWSPFGPIGAVIINTVTVSLSAKVSCQKHNYG	2646
QY	2555	KKGVSGLQSPFAVLLLSATWMLLALSVNSDPLLFFYLFPATCNCIOGPEIFLSVYLISK	2614
Db	2647	KKGIVSLRLRPAFLILLIISATWMLGLLAVNRDALSFFYTLFALISGLQGFVLLPHCYLNG	2706
QY	2615	EVRKALK-LACSKRPSPDPLTTKST-LTSSVNCPSPYADG--RLYOPYGDASGSLHSTS	2670
Db	2707	EVRKMLKGVILGGRKLMLEDATTRATLLTSLMNTTFFGCGPDMRLTDLBSTASLDSIV	2766
QY	2671	RS-----GKSQPSYVFLRLRESBALNPGQGPRLGDLPGSLFLREGDQD	2713
Db	2767	RDEGICQLGTVSSGLVNGSHGEPDASLMPRCXD-----PPG-----	2802
QY	2714	HDPDTSDSLSTBEDDQSGSYASTHSSDSEEBEERBEAAPFGBOQMSILGFGAERLP	2773
Db	2803	--HDSDBDSHSLSL-DEQSSSYAASHSDSDSDDCGVAERK-----WDPARG-----A	2845
QY	2774	LHSPNK-DGSPRGGRKAPWPGDPFTTAKESSGGNAP-----BERLRENGDAL	2818
Db	2846	VHSTPDKDAAVANHVPAWMPDQSLAESSESDPSGCPRLKVTETKYSVELHREBQSGHREYR	2905
QY	2819	SREBSLGLPFGSSAOP-----HKGILLKKKCL-----PTISEKS--SILRLPLBOCTGSSRGSS	2869
Db	2906	PQDSGGGAARLASQPEBQKGIILKOKKVTYPPPLTLITBOQLKGLRLAKRLADCEGSPSSR	2965
QY	2870	ASBBSRGPF-----PRPFRPQOSLOBOQLNCTMPPLAMSTKAGTVDEBDSGSE	2915
Db	2966	TSLSIGSGGPICAITVKSPPGREPRDHLNAGV---AMNVRTSSAADGSDSE	3012

Search completed: February 11, 2004, 16:05:02  
Job time : 79 secs

Db 2107 ISFVDLRAMNEKLSRNETQVDGARALQLVRLRSATQHTGTLEGNDRVATAYQLLGHVLQH 2166

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 11, 2004, 15:50:43 ; Search time 61 Seconds  
(without alignments)  
7605.865 Million cell updates/sec

Title: US-09-916-849a-3

Perfect score: 15545  
Sequence: 1 MRSPTGVLPPTPPPLLL.....AGTVDESSGSEFLFNFIAH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 156726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq.130un03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15545	100.0	2923	22	AAU07054 Human Flamingo pro
2	15545	100.0	2923	23	AAW50866 Cadherin EGF lag s
3	15545	100.0	2923	24	ABP82018 Human GPCR CELSR2
4	15518.5	99.8	2956	22	AAU07053 Human Flamingo pol
5	15393	99.0	2894	24	ABU11556 Human MDP1 polypep
6	15279	98.3	2936	23	AAU74826 Human REPT9 prot
7	13399	86.2	2560	22	ABU11404 Human FLAMINGO 1 h
8	12541.5	80.7	2405	21	AAW42192 Human ORFX ORF1956
9	8754	56.3	3014	22	AAW65533 Human novel cytocki

10	8754	56.3	3014	22	AAU02196	Seven-pass transme
11	8754	56.3	3014	22	ABP81979	Human GPCR CELSR1/
12	8709	56.0	3028	22	AAW08586	Human NOV7 protein
13	8495	54.6	2707	18	AAW27161	Mouse receptor ME2
14	7816.5	50.3	4115	23	ABP53570	Human NOV2 protein
15	7763	49.9	3298	22	AAW03657	Human extracellular
16	4817.5	31.0	3606	22	ABW62595	Drosophila melanog
17	4369.5	28.1	1572	18	AAW27160	Mouse receptor ME2
18	4093	26.3	1460	24	ABP81741	Human CELSR3 prote
19	3641	23.4	717	21	AAW56721	Human prostate can
20	2958	19.0	568	19	AAW51244	Human calcitonin r
21	2937	18.9	568	23	ABW89665	Human polypeptide
22	2900.5	18.7	565	24	ABJ37074	Human breast cance
23	2425.5	15.6	1554	22	ABW11890	Human protocadheri
24	1814.5	11.7	470	24	ABP75877	Human secretory po
25	1712.5	11.0	4903	24	ABU12072	Human NOV17a CG928
26	1534.5	9.9	5147	22	ABW59831	Drosophila melanog
27	1516.5	9.8	4590	23	ABW68608	Human pancreatic c
28	1516.5	9.8	4590	24	AAW26790	Human cadherin (CA
29	1516.5	9.8	4590	24	AAW26791	Human cadherin (CA
30	1516.5	9.8	4591	22	ABW22977	Novel human diagno
31	1477.5	9.5	4643	22	ABW11609	Drosophila melanog
32	1419.5	9.1	3097	22	ABW62967	Drosophila melanog
33	1413	9.1	4555	23	AAW52106	Rat fat 3 protein
34	1336.5	8.6	3909	23	ABP43661	Human protocadheri
35	1320.5	8.5	4349	23	AAU79940	Novel human protei
36	1320.5	8.5	4349	23	ABW97540	Novel human protei
37	1320.5	8.5	4349	24	AAW26792	Human cadherin (CA
38	1262.5	8.1	3298	23	ABW05430	Human dachshous pro
39	1262.5	8.1	3298	24	AAW26793	Human cadherin (CA
40	1260.5	8.1	4263	23	ABW97541	Novel human protei
41	1217.5	7.8	3354	24	AAW26788	Human cadherin (CA
42	1217.5	7.8	3354	24	AAW26789	Human cadherin (CA
43	1210.5	7.8	3217	23	AAU74825	Human REPT9 prot
44	1207	7.8	3503	22	ABW64499	Drosophila melanog
45	1137.5	7.3	1783	22	ABW63321	Drosophila melanog

#### ALIGNMENTS

RESULT 1	AAU07054	standard; Protein; 2923 AA.
ID	AAU07054	standard; Protein; 2923 AA.
AC	AAU07054;	
XX		
DT	24-OCT-2001	(first entry)
XX		
DE	Human Flamingo protein encoded by cDNA splice variant.	
XX		
KW	Flamingo; human; splice variant; G-protein coupled receptor; diabetes;	
KW	signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;	
KW	anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;	
KW	obesity; hypotension; hypertension; urinary retention; angina pectoris;	
KW	myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;	
KW	benign prostatic hyper trophy; psychotic disorder; neurological disorder;	
KW	manic depression; delirium; dementia; severe mental retardation;	
KW	Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;	
KW	antifungal; antiviral; antiprotocadherin; anti-HIV; anorectic; anti-anorectic;	
KW	antiparkinsonian; cardiant; cerebroprotective; neuroprotective;	
KW	antidepressant; anticonvulsant; antisense therapy; gene therapy.	
OS	Homo sapiens.	
XX		
PN	WO200161003-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	19-FEB-2001; 2001WO-GB00680.	
XX		
PR	19-FEB-2000; 2000GB-0004196.	
XX		

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Testa TT;  
 XX  
 DR WPI; 2001-502792/55.  
 DR N-PSDB; AAS11678.  
 XX  
 PT An isolated Flamingo polypeptide useful for treating diseases such as  
 PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 PT osteoporosis -  
 PS  
 PS Claim 2; Page 33-34; 66pp; English.  
 XX  
 XX The sequence represents a human Flamingo polypeptide encoded by a cDNA  
 CC splice variant polynucleotide. Flamingo is a member of the G-protein  
 CC coupled receptor family, which is involved in signal transduction  
 CC pathways. By screening to identify compounds that stimulate or inhibit  
 CC the function or level of the protein, treatments can be developed for  
 CC various diseases and bacterial, fungal, protozoan and viral infections,  
 CC including HIV, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, angina pectoris, myocardial infarction, stroke,  
 CC ulcers, allergies and benign prostatic hypertrophy. Also treatable are  
 CC psychotic and neurological disorders such as anxiety, schizophrenia,  
 CC manic depression, delirium, dementia, severe mental retardation,  
 CC Huntington's disease and Gilles de la Tourette's syndrome.  
 CC  
 XX  
 XX Sequence 2923 AA;  
 SQ  
 Query Match 100.0%; Score 15545; DB 22; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSPATGVPPTPPPLLLLLLLLLLPPPLGGDVGPRSLGSRGSGSCAPMGMLCPB 60  
 DB 1 MRSPATGVPPTPPPLLLLLLLLLLPPPLGGDVGPRSLGSRGSGSCAPMGMLCPB 60  
 QY 61 SASNLMLYTSRCDAGTELTHGLVPHHDGLVWCPSSEAHIPLPAPBEGCPMGCRLIGIG 120  
 DB 61 SASNLMLYTSRCDAGTELTHGLVPHHDGLVWCPSSEAHIPLPAPBEGCPMGCRLIGIG 120  
 QY 121 GHSPQCKLTLPBHPCLKAPRLRCOSCKLAOPGLRAGERSPEBSIGRRKNVTAPO 180  
 DB 121 GHSPQCKLTLPBHPCLKAPRLRCOSCKLAOPGLRAGERSPEBSIGRRKNVTAPO 180  
 QY 121 GHSPQCKLTLPBHPCLKAPRLRCOSCKLAOPGLRAGERSPEBSIGRRKNVTAPO 180  
 DB 121 GHSPQCKLTLPBHPCLKAPRLRCOSCKLAOPGLRAGERSPEBSIGRRKNVTAPO 180  
 QY 181 FQPPSYOATVPENOPACTPVASLRAPIDPDEGEAGRLTYMDALPDSNSQFSLDPYGA 240  
 DB 181 FQPPSYOATVPENOPACTPVASLRAPIDPDEGEAGRLTYMDALPDSNSQFSLDPYGA 240  
 QY 241 VTTABELDRBTKSTHYFRVTAODHGMFRSALATLTLLVTDNDHPVPEQOBYKSLRE 300  
 DB 241 VTTABELDRBTKSTHYFRVTAODHGMFRSALATLTLLVTDNDHPVPEQOBYKSLRE 300  
 QY 301 NLEVGIVLVTRATDGAAPRANILYLLRSGSGSPSEVEIDPRSGVITRRPVVREEV 360  
 DB 301 NLEVGIVLVTRATDGAAPRANILYLLRSGSGSPSEVEIDPRSGVITRRPVVREEV 360  
 QY 361 ESYOVLTEASDOGRDPGRSTTAAPFLSVEDDNDNAQFSEKRYVVOVREDYTPGAPVLR 420  
 DB 361 ESYOVLTEASDOGRDPGRSTTAAPFLSVEDDNDNAQFSEKRYVVOVREDYTPGAPVLR 420  
 QY 421 VTTASDRKGSNAVVHYSIMSGNAGQFYLLDAQTGALDVSPLDYETTKXYTLNVAQDGG 480  
 DB 421 VTTASDRKGSNAVVHYSIMSGNAGQFYLLDAQTGALDVSPLDYETTKXYTLNVAQDGG 480  
 QY 481 RPLPSNVSLVTVOVLDINNAPIFVSTPPOATVLESVPLGYLVTHQAIDAAAGNARL 540  
 DB 481 RPLPSNVSLVTVOVLDINNAPIFVSTPPOATVLESVPLGYLVTHQAIDAAAGNARL 540  
 QY 541 EYRLAGVGHDPFPTINNGTGWISVAELDREEDVFSFGVEARDHGTALTAASVSVTV 600  
 DB 541 EYRLAGVGHDPFPTINNGTGWISVAELDREEDVFSFGVEARDHGTALTAASVSVTV 600

QY 601 LDVNDNNPTFQPEXTVNLNEDAAVGSVTVSAVDRDASHVITVQITSGTRNRKFSITS 660  
 DB 601 LDVNDNNPTFQPEXTVNLNEDAAVGSVTVSAVDRDASHVITVQITSGTRNRKFSITS 660  
 QY 661 QSGGGLVSLAPLDYKLERQYVLAVTASDGTRODTAQIVNVTTDANTRPVFOSSHVTVN 720  
 DB 661 QSGGGLVSLAPLDYKLERQYVLAVTASDGTRODTAQIVNVTTDANTRPVFOSSHVTVN 720  
 QY 721 VVEDRPAGTVVLSATDEDEGENARITYFMEDSIPQRIADTGAVTQOAELEDVDOVS 780  
 DB 721 VVEDRPAGTVVLSATDEDEGENARITYFMEDSIPQRIADTGAVTQOAELEDVDOVS 780  
 QY 781 YTLATTAARDNGIPKSDTTYLEILVNDVNDNAPOELRDSYOGSYVEDVPFTSVLQISAT 840  
 DB 781 YTLATTAARDNGIPKSDTTYLEILVNDVNDNAPOELRDSYOGSYVEDVPFTSVLQISAT 840  
 QY 841 DSDSLNGRVFTTFCGGDGGDGFVVESTSGVTRLRDLRENNVQYTLRAYADKGMPP 900  
 DB 841 DSDSLNGRVFTTFCGGDGGDGFVVESTSGVTRLRDLRENNVQYTLRAYADKGMPP 900  
 QY 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVTAIDPDEGTNAQIMYQI 960  
 DB 901 ARTPMETVTVLDVNDNPPVFEQDEDFVFEENSPIGLAVARVTAIDPDEGTNAQIMYQI 960  
 QY 961 VEGNIPEVFOULDFSGELTALVDLDYEDRPEYVVIQATSAELVSATVAVNLLDRNDP 1020  
 DB 961 VEGNIPEVFOULDFSGELTALVDLDYEDRPEYVVIQATSAELVSATVAVNLLDRNDP 1020  
 QY 1021 PVLGNFELLFNNTYNNRSSSPFGAIGRVPAHDPIISLSTYSFRGNELSVLLNASTG 1080  
 DB 1021 PVLGNFELLFNNTYNNRSSSPFGAIGRVPAHDPIISLSTYSFRGNELSVLLNASTG 1080  
 QY 1141 FLSPILGLFTQVAATLTATPPDHVVVFNVRDTPADGSHILNLSVQPPGCGPPFL 1200  
 DB 1141 FLSPILGLFTQVAATLTATPPDHVVVFNVRDTPADGSHILNLSVQPPGCGPPFL 1200  
 QY 1201 PSEDLOERLYNRSLLTASORVLPDPDNTCLREPCNMYMCVVLAPDSGAPFIASS 1260  
 DB 1201 PSEDLOERLYNRSLLTASORVLPDPDNTCLREPCNMYMCVVLAPDSGAPFIASS 1260  
 QY 1261 VLFRRPIHPVGLRCRCPGFTGDCYETEVDCYSRPGCHGRSREGGYTCLCRDGYTG 1320  
 DB 1261 VLFRRPIHPVGLRCRCPGFTGDCYETEVDCYSRPGCHGRSREGGYTCLCRDGYTG 1320  
 QY 1321 EHCYVSANSGRCTPGVCNKGTCVNLVGGFKDCPCSGDPFEXYCOVTTSRSPAFSFTF 1380  
 DB 1321 EHCYVSANSGRCTPGVCNKGTCVNLVGGFKDCPCSGDPFEXYCOVTTSRSPAFSFTF 1380  
 QY 1381 RGLRORFHTLALSAATBRDGLLLYNGRFNKKHDFVLEVLQEOVLTFSGESTTYS 1440  
 DB 1381 RGLRORFHTLALSAATBRDGLLLYNGRFNKKHDFVLEVLQEOVLTFSGESTTYS 1440  
 QY 1441 PVPVGSVSDGQWHTYOLKTYNRPLLGOTGLPOGSEKQKAVVTVYDSCDTGVALRGSYLG 1500  
 DB 1441 PVPVGSVSDGQWHTYOLKTYNRPLLGOTGLPOGSEKQKAVVTVYDSCDTGVALRGSYLG 1500  
 QY 1501 NYSCAAQGTQGGSKSLDTGELLGCVPDLPBSPPVNRQROVGCGRMLQVDSRIIDMAD 1560  
 DB 1501 NYSCAAQGTQGGSKSLDTGELLGCVPDLPBSPPVNRQROVGCGRMLQVDSRIIDMAD 1560  
 QY 1561 FIANNGTVPACPAKXNVCDSNTCHNGGTCVNMOWDAFSCBCPLGFGGSKCAQEMANPQHL 1620  
 DB 1561 FIANNGTVPACPAKXNVCDSNTCHNGGTCVNMOWDAFSCBCPLGFGGSKCAQEMANPQHL 1620  
 QY 1621 GSSLVAMHGLSLPISQPMYLSIMFRTQADGVLLQATIRGRSTITLQAREGHVMLSVEGT 1680  
 DB 1621 GSSLVAMHGLSLPISQPMYLSIMFRTQADGVLLQATIRGRSTITLQAREGHVMLSVEGT 1680  
 QY 1681 GLQASRLRLEPRANDGDMHNAQLALGASGCGHAIISLSDYQQAEGNLGRLGHLHS 1740

```

Db 1661 GLOASLLEPERADGWHHQAQLGASGGGHAHLISFDYQOQRAEENGLRGLHGLHS 1740
Qy 1741 NITVGGIIPGAPGAVARGFRCIQGVRSVDPTEGVNSLDPSHGESINVEQCSLPDPCSN 1800
Db 1741 NITVGGIIPGAPGAVARGFRCIQGVRSVDPTEGVNSLDPSHGESINVEQCSLPDPCSN 1800
Qy 1801 PCPANSYCSNMDYSYCSGCDPEYTDNCTNVCNLPCEHQSVCYTRKPSAPHGYTCPCPN 1860
Db 1801 PCPANSYCSNMDYSYCSGCDPEYTDNCTNVCNLPCEHQSVCYTRKPSAPHGYTCPCPN 1860
Qy 1861 YLGPYCETRIIDPCRGMMGHPTGCPNCDVSKGPDPCNKTSGCHCKENHYPGSPPT 1920
Db 1861 YLGPYCETRIIDPCRGMMGHPTGCPNCDVSKGPDPCNKTSGCHCKENHYPGSPPT 1920
Qy 1921 CLLCDYPTGSLSRVCDPEDCQCPCKPGVIGRCDCRCNPNFAEVTYNGCEVYDSCPRAI 1980
Db 1921 CLLCDYPTGSLSRVCDPEDCQCPCKPGVIGRCDCRCNPNFAEVTYNGCEVYDSCPRAI 1980
Qy 1981 EAGIWPPTRRGLPAAPCPKGSFGTAVRHCDHRGMLPMLFNCTSIITFSELKGPAREL 2040
Db 1981 EAGIWPPTRRGLPAAPCPKGSFGTAVRHCDHRGMLPMLFNCTSIITFSELKGPAREL 2040
Qy 2041 QRNESGLDSGRSQOALILPNAHTAGYFGSDVAVAYQALTRILAHBSTORGFGLSATQ 2100
Db 2041 QRNESGLDSGRSQOALILPNAHTAGYFGSDVAVAYQALTRILAHBSTORGFGLSATQ 2100
Qy 2101 DVHFTENILRVGSALLDPTANKRMELIOQTEGTMLOHYEAAYSALAONMHTYLSPF 2160
Db 2101 DVHFTENILRVGSALLDPTANKRMELIOQTEGTMLOHYEAAYSALAONMHTYLSPF 2160
Qy 2161 TIVTBNIVISVVRLDKGNPAGAKLPRYBALNGEOPDLETTVIIPESVFRETPEVVRAG 2220
Db 2161 TIVTBNIVISVVRLDKGNPAGAKLPRYBALNGEOPDLETTVIIPESVFRETPEVVRAG 2220
Qy 2221 PGEAOPEBELARBORHPELSQGEAVASVITITRGLGLPHNYOPDKSLAVPRPIINT 2280
Db 2221 PGEAOPEBELARBORHPELSQGEAVASVITITRGLGLPHNYOPDKSLAVPRPIINT 2280
Qy 2281 FVVISIVHDEBELLPALDKPVTVQFRLLETERKPCIVFMNHSILVSGTGSARCE 2340
Db 2281 FVVISIVHDEBELLPALDKPVTVQFRLLETERKPCIVFMNHSILVSGTGSARCE 2340
Qy 2341 VVFRNESVSCQCNMTSPAVLMDVSRRENGEILPLKLTITVALGVTLAALLTFEFTL 2400
Db 2341 VVFRNESVSCQCNMTSPAVLMDVSRRENGEILPLKLTITVALGVTLAALLTFEFTL 2400
Qy 2401 LRIIRSNHGIRBNLTALGLAQLVFLGINOADI.PFACVIAIILHFLYICTFSWALL 2460
Db 2401 LRIIRSNHGIRBNLTALGLAQLVFLGINOADI.PFACVIAIILHFLYICTFSWALL 2460
Qy 2461 ALHLYRALTEVDVNTGPMRFYMLGMGVPAFITGLAVGLDPBGVGNDFCMLSYDTLI 2520
Db 2461 ALHLYRALTEVDVNTGPMRFYMLGMGVPAFITGLAVGLDPBGVGNDFCMLSYDTLI 2520
Qy 2521 WSPFAGPVAFAVMSVFLYITLARAASCAAQROGFEXKGVSGIQSFVALLLSATWLLAL 2580
Db 2521 WSPFAGPVAFAVMSVFLYITLARAASCAAQROGFEXKGVSGIQSFVALLLSATWLLAL 2580
Qy 2581 LSVNSDTLLFHYLPATCNCIGPFIPLSYVVLSTKEVRALCLAGSRKSPDPAITXTSL 2640
Db 2581 LSVNSDTLLFHYLPATCNCIGPFIPLSYVVLSTKEVRALCLAGSRKSPDPAITXTSL 2640
Qy 2641 TTSYNCBPYADGRLYQPYGDSAGLSHSTRSGSQPSYIFPLLREESALNPGQPGILG 2700
Db 2641 TTSYNCBPYADGRLYQPYGDSAGLSHSTRSGSQPSYIFPLLREESALNPGQPGILG 2700
Qy 2701 DPGSIFLEGQOQHPDPTSDSLSDBDQSGASTHSSSSEEEEEEBAAPGEOG 2760
Db 2701 DPGSIFLEGQOQHPDPTSDSLSDBDQSGASTHSSSSEEEEEEBAAPGEOG 2760
Qy 2761 WDSLILGPAERLPLHSTPKDGGPGKAPWPGDFTTAKESSGNKAPERRRENGDALSR 2820

```

```

Db 2761 WDSLILGPAERLPLHSTPKDGGPGKAPWPGDFTTAKESSGNKAPERRRENGDALSR 2820
Qy 2821 EGSIGPLPGSSAOPHKILKKKCLPTISSEKSLRLPLBOCTGSSRGSASGSGGPP 2880
Db 2821 EGSIGPLPGSSAOPHKILKKKCLPTISSEKSLRLPLBOCTGSSRGSASGSGGPP 2880
Qy 2881 RPPRQSLQEQNLQWPIIAMSIRKAGTVDEDSGSEPLFFNFLH 2923
Db 2881 RPPRQSLQEQNLQWPIIAMSIRKAGTVDEDSGSEPLFFNFLH 2923

RESULT 2
AAM50866
ID AAM50866 standard; Protein; 2923 AA.
AC AAM50866;
DT 07-MAY-2002 (first entry)
DE Cadherin EGF LAG seven-pass G-type receptor 2, basal cell marker.
KW Cadherin EGF LAG seven-pass G-type receptor 2; receptor;
KW EGF-like domain, multiple 2; basal cell; marker; breast cancer;
KW tumour; diagnosis; prognosis; classification; human; gene therapy.
OS Homo sapiens.
PN MOJ00208765-A2.
PD 31-JAN-2002.
PE 26-JUL-2001; 2001MO-US23843.
PR 26-JUL-2000; 2000US-220967P.
XX (STRD ) UNIV STANFORD.
PA (GENO-) APPLIED GENOMICS INC.
PI Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
PI Seltz R;
XX WPI; 2002-206094/26.
XX
XX Classifying and treating breast tumours associated with the expression
XX and activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14
XX and/or cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like
XX domain, Multiple 2 -
XX
XX Claim 3, Fig 1C; 741pp; English.
XX
XX The present sequence is that of human cadherin EGF LAG seven-pass
XX G-type receptor 2, also known as EGF-like domain, multiple 2.
XX The invention provides new reagents and methods for the management
XX (e.g. detection, classification, provision of diagnostic and
XX prognostic information, treatment, etc.) of breast cancer. cDNA
XX microarray technology was used to identify genes whose expression
XX profile across a large group of tumour samples correlated with that
XX of cytokeratin 5 and cytokeratin 17, markers for basal cells of the
XX normal mammary lactation gland. Tumours that express cytokeratin
XX 5/6 and/or 17 have a poor prognosis relative to tumours overall.
XX The basal marker genes identified included those encoding cadherin
XX 3 (see AAM50864), matrix metalloproteinase 14 (see AAM50865) and
XX cadherin EGF LAG seven-pass G-type receptor 2. Detection of these
XX markers allows classification of a tumour sample as belonging to a
XX basal tumour subclass. By providing reagents that may reliably be
XX used to classify tumours as belonging to a basal subclass, the
XX invention enables a variety of methods for improving therapeutic
XX options for patients with breast cancer and for individualising
XX therapy. Potential therapeutic agents include compounds that
XX modulate breast basal cell marker genes or that modulate
XX polypeptides encoded by these genes. In particular, therapeutic
XX agents include antibodies that specifically bind to the basal
XX marker proteins and which recognise basal cells of normal mammary

```



CC laccation glands. The antibodies can be used either by themselves  
CC or when conjugated to, or delivered with, another molecule such as  
CC a toxic compound. The likelihood that cadherin EGF lig seven-pass  
CC G-type receptor 2 is membrane-bound makes it an attractive candidate  
CC for antibody therapeutics.

XX Sequence 2923 AA:

Query Match 100.0%; Score 15545; DB 23; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSRATGVPLEPPPLLLLLLLPPPLLDGVGCRSISGRSSGACAPMGLCPSS 60  
DB 1 MRSRATGVPLEPPPLLLLLLLPPPLLDGVGCRSISGRSSGACAPMGLCPSS 60  
QY 61 SASNMLYTSRCDAGTELTGHLVPHHDGLRWCPSESAHPLPPAPGCPMSCLLGIG 120  
DB 61 SASNMLYTSRCDAGTELTGHLVPHHDGLRWCPSESAHPLPPAPGCPMSCLLGIG 120  
QY 121 GHLSPOGKLTLPREHCLKAPRLRCSCKLAQAPGAPGERSPEBSIGGRKXNVTAPO 180  
DB 121 GHLSPOGKLTLPREHCLKAPRLRCSCKLAQAPGAPGERSPEBSIGGRKXNVTAPO 180  
QY 181 POPSYQATVPENQAPGTPVASLRALDPDEGAERLEYTMDALFDSRNOFSLDPVTA 240  
DB 181 POPSYQATVPENQAPGTPVASLRALDPDEGAERLEYTMDALFDSRNOFSLDPVTA 240  
QY 241 VTTAEELDEETKSTHYFRYTAQDHGMPRRSALATLTTLVTDTMDHPVPEQCKYSLAR 300  
DB 241 VTTAEELDEETKSTHYFRYTAQDHGMPRRSALATLTTLVTDTMDHPVPEQCKYSLAR 300  
QY 301 NLEVGVEVLTARITDAPPNANILYRLLEGSGSPSEVEIDPRSGVIRTRGVNREXY 360  
DB 301 NLEVGVEVLTARITDAPPNANILYRLLEGSGSPSEVEIDPRSGVIRTRGVNREXY 360  
QY 361 ESYQVLTVEASDQGRDPGRSTTAAPVLSVEDNDNAPQSEKRYVYVQVEDVTPGAPVLR 420  
DB 361 ESYQVLTVEASDQGRDPGRSTTAAPVLSVEDNDNAPQSEKRYVYVQVEDVTPGAPVLR 420  
QY 421 VTSASDDKSNAAVHYSIMSGNARGOFTYLDQAGLADVSPDYETTKETTLKVRQDDG 480  
DB 421 VTSASDDKSNAAVHYSIMSGNARGOFTYLDQAGLADVSPDYETTKETTLKVRQDDG 480  
QY 481 RPLSLNVSGLVTVQVLDINDNAPIFVSTPQATVLESPLGYLVHQAIDADAGNARL 540  
DB 481 RPLSLNVSGLVTVQVLDINDNAPIFVSTPQATVLESPLGYLVHQAIDADAGNARL 540  
QY 541 EYRLAGVGHDPPTINNCTGMSVAELDREEDVPSFGVEARDHGTPALTASASVTV 600  
DB 541 EYRLAGVGHDPPTINNCTGMSVAELDREEDVPSFGVEARDHGTPALTASASVTV 600  
QY 601 LDVNDNNPFTQPEBYTRLNEDAAVGSVTVSAVNDASHVITYOITSGNTNRSITS 660  
DB 601 LDVNDNNPFTQPEBYTRLNEDAAVGSVTVSAVNDASHVITYOITSGNTNRSITS 660  
QY 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVVTDMANTHRFVPOSHYTN 720  
DB 661 QSGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAQIVNVVTDMANTHRFVPOSHYTN 720  
QY 721 VNEBPRAGITVVLISATDEBTGENARITYEMEDSIPOFRIDADTGAVTTOABLDYEDQVS 780  
DB 721 VNEBPRAGITVVLISATDEBTGENARITYEMEDSIPOFRIDADTGAVTTOABLDYEDQVS 780  
QY 781 YTLAATRRDNGIPQKSTYTLLELVNDVNDNAPQFLDSYQGSYVEBVPFTSVLQISAT 840  
DB 781 YTLAATRRDNGIPQKSTYTLLELVNDVNDNAPQFLDSYQGSYVEBVPFTSVLQISAT 840  
QY 841 DRDSGLNGRVYTYFGGDDGDFIVESTSGIVTLERLDRBNVAQYVLAAYVNDKMP 900  
DB 841 DRDSGLNGRVYTYFGGDDGDFIVESTSGIVTLERLDRBNVAQYVLAAYVNDKMP 900  
QY 901 ARTPEMTVTVLVLDVNDNPPVPEQDEDFVPEBNSPIGLAARVTAIDPDGCTNAQIMYQI 960

DB 901 ARTPEMTVTVLVLDVNDNPPVPEQDEDFVPEBNSPIGLAARVTAIDPDGCTNAQIMYQI 960  
QY 961 VEGNIPEVPOLDIFSGELTALVDLYEDRPREVVIYQNTSAPLSARATVHRLDRNMP 1020  
DB 961 VEGNIPEVPOLDIFSGELTALVDLYEDRPREVVIYQNTSAPLSARATVHRLDRNMP 1020  
QY 1021 PVLGNFELLFNNYVYNNRSSPFGAIGRVPAHDPDISLTYSPFRGNELSVLNLASTG 1080  
DB 1021 PVLGNFELLFNNYVYNNRSSPFGAIGRVPAHDPDISLTYSPFRGNELSVLNLASTG 1080  
QY 1081 EIKLSRALDNNRPLEAIVSVSDGSHVTAQALRVITIIDEMLTHTSITTLRLBMSPER 1140  
DB 1081 EIKLSRALDNNRPLEAIVSVSDGSHVTAQALRVITIIDEMLTHTSITTLRLBMSPER 1140  
QY 1141 FLSPILGLFIOAVATLTATPPDHVVVFNVDTPARAGHILNVSLVSGOPPGPGGAPPL 1200  
DB 1141 FLSPILGLFIOAVATLTATPPDHVVVFNVDTPARAGHILNVSLVSGOPPGPGGAPPL 1200  
QY 1201 PSEDLOERLYNRSLLTAISAQRVLPDDNICTAREPCENYMCVSLNFDSSAPFIASSS 1260  
DB 1201 PSEDLOERLYNRSLLTAISAQRVLPDDNICTAREPCENYMCVSLNFDSSAPFIASSS 1260  
QY 1261 VLFRIHVGGLRCRCPGFTGDIYCTEVDLCYSRPGCPHRCRSRGGYTLCRDGYTG 1320  
DB 1261 VLFRIHVGGLRCRCPGFTGDIYCTEVDLCYSRPGCPHRCRSRGGYTLCRDGYTG 1320  
QY 1321 EHCYSASGRCTPVCNNGGVNVLVYGFPCDPSGDFEPCVQVTRSPASFTF 1380  
DB 1321 EHCYSASGRCTPVCNNGGVNVLVYGFPCDPSGDFEPCVQVTRSPASFTF 1380  
QY 1381 RGLRQRFHTLALSPATEKRDGLLYNGRPNEKHDVALLEYLOEVOVLTFSAGSESTTVS 1440  
DB 1381 RGLRQRFHTLALSPATEKRDGLLYNGRPNEKHDVALLEYLOEVOVLTFSAGSESTTVS 1440  
QY 1441 PVPBGVSDGQMHYTLQKYNKPLIGOTGLPOGSEBOYAVYTVDCGCTGVALRGSYLG 1500  
DB 1441 PVPBGVSDGQMHYTLQKYNKPLIGOTGLPOGSEBOYAVYTVDCGCTGVALRGSYLG 1500  
QY 1501 NYSCAAQGTGGGSKSLDTPGLLGGVBDLPESPPVMMRQPVGCMRLNQVDSRIIDMAD 1560  
DB 1501 NYSCAAQGTGGGSKSLDTPGLLGGVBDLPESPPVMMRQPVGCMRLNQVDSRIIDMAD 1560  
QY 1561 FIANNGTVPQCPAKKNVDSNTCHNGGTCVNMQDAPSCBCLPFGSKSCAQOEMANPOHFL 1620  
DB 1561 FIANNGTVPQCPAKKNVDSNTCHNGGTCVNMQDAPSCBCLPFGSKSCAQOEMANPOHFL 1620  
QY 1621 GSSLVAMHGLSLPIQPMYTLIMFRTROADGVLQAITRGRTITLQLRBGRVMSVSGT 1680  
DB 1621 GSSLVAMHGLSLPIQPMYTLIMFRTROADGVLQAITRGRTITLQLRBGRVMSVSGT 1680  
QY 1681 GLQASSLRLEPRANDGMHHAQLALGASGGGHAHLSFDVYQOARAGNLGRLGLHLS 1740  
DB 1681 GLQASSLRLEPRANDGMHHAQLALGASGGGHAHLSFDVYQOARAGNLGRLGLHLS 1740  
QY 1741 NITVGGIPGAPGAVARGCLQGVRSVDPBGVNSLDPBSHESINVEQCSGLPDPCCSN 1800  
DB 1741 NITVGGIPGAPGAVARGCLQGVRSVDPBGVNSLDPBSHESINVEQCSGLPDPCCSN 1800  
QY 1801 PCPANSYSCNMDVSYSCDPPGYDNCNVDLNPCHQSCTKRPSPAPHYTCBCEPN 1860  
DB 1801 PCPANSYSCNMDVSYSCDPPGYDNCNVDLNPCHQSCTKRPSPAPHYTCBCEPN 1860  
QY 1861 YLGPYCESTRIDQPCRGWGHPTGCPKCNVDSKGFDPDCKNTSGRCHCKENYRPPGSP 1920  
DB 1861 YLGPYCESTRIDQPCRGWGHPTGCPKCNVDSKGFDPDCKNTSGRCHCKENYRPPGSP 1920  
QY 1921 CLLCDCTPTGSLSRVCDPEDGQCPCKPGVYTGQCDRCNDPABVYTNNGCVYDSCPRAI 1980  
DB 1921 CLLCDCTPTGSLSRVCDPEDGQCPCKPGVYTGQCDRCNDPABVYTNNGCVYDSCPRAI 1980  
QY 1981 BAGIWMPTRRFGLPAAAPCPKSGPATAVHCDENHGMPLPNNFNCTSTFSELKGPABEL 2040

```

Db      1981  EAGIMWRTRFGLPAAPCPKSGFCAVHCHDHGMLPNNLNCSTSTPSEIKGAEKL 2040
Qy      2041  QNNESGLDGRSQOALLRNATQTAGYFGSDVKAAYOLATLLAHSTORFGLSAQO 2100
Db      2041  QNNESGLDGRSQOALLRNATQTAGYFGSDVKAAYOLATLLAHSTORFGLSAQO 2100
Qy      2101  DVHFTNLRVSGALLDTANKRMWELIQOEGTALLOHYEYASALANNNHTYLSPP 2160
Db      2101  DVHFTNLRVSGALLDTANKRMWELIQOEGTALLOHYEYASALANNNHTYLSPP 2160
Qy      2161  TTVTPNIVISVVRDLKGNFAGAKLPRYBALRGEQPDLETTVILPSSVRETPPVVRPAG 2220
Db      2161  TTVTPNIVISVVRDLKGNFAGAKLPRYBALRGEQPDLETTVILPSSVRETPPVVRPAG 2220
Qy      2221  PGEAOPPEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLRVPKPEIINT 2280
Db      2221  PGEAOPPEELARORRHPELSQGEAVASYIYRTLAGLLPHNYDPDKSLRVPKPEIINT 2280
Qy      2281  PVSISVHDEBELPALDKPVTVQFRLLETERKPCICVFNHSLIVSGTGMASRGCE 2340
Db      2281  PVSISVHDEBELPALDKPVTVQFRLLETERKPCICVFNHSLIVSGTGMASRGCE 2340
Qy      2341  VVFRNESHVSCCNHMTSPAVLMDVSRRENGELLPLKTLTYVALGVTLAALLTPPELT 2400
Db      2341  VVFRNESHVSCCNHMTSPAVLMDVSRRENGELLPLKTLTYVALGVTLAALLTPPELT 2400
Qy      2401  LRILSNQHGIRRNLTALALQVVELGIGNADLPACTVIALHLFYLCTFSFALLE 2460
Db      2401  LRILSNQHGIRRNLTALALQVVELGIGNADLPACTVIALHLFYLCTFSFALLE 2460
Qy      2461  AHLKRALTRVDVNTGPMRFYMGVPAFTTGLAVLDEBGNGNPPCMLSTYDTLI 2520
Db      2461  AHLKRALTRVDVNTGPMRFYMGVPAFTTGLAVLDEBGNGNPPCMLSTYDTLI 2520
Qy      2521  WSPFAGVAVASVVELYTLAARASCAQROGFEKKGPVSGQPSFAYILLSATMLAL 2580
Db      2521  WSPFAGVAVASVVELYTLAARASCAQROGFEKKGPVSGQPSFAYILLSATMLAL 2580
Qy      2581  LSVNSDTLFFHYLFATCNCIQGPFIPLSYVLSKVRKALKACSRKSPDPAITTKSTL 2640
Db      2581  LSVNSDTLFFHYLFATCNCIQGPFIPLSYVLSKVRKALKACSRKSPDPAITTKSTL 2640
Qy      2641  TSSVNCPSPYADGRLYOPYGDAGSLHSTRSGKOPSTIPLLEESALNPGQPPGLG 2700
Db      2641  TSSVNCPSPYADGRLYOPYGDAGSLHSTRSGKOPSTIPLLEESALNPGQPPGLG 2700
Qy      2701  DPGSLFLBQDQOHPDTSDDSLLEDDQSGSYASTHSSDBEEREBEEREAAPPGGOG 2760
Db      2701  DPGSLFLBQDQOHPDTSDDSLLEDDQSGSYASTHSSDBEEREBEEREAAPPGGOG 2760
Qy      2761  WDSLIGPQAEKRLPLHSTPKDGGPGFGKAPWPGDFTTAKSSGNGAPBEERLENGDALSR 2820
Db      2761  WDSLIGPQAEKRLPLHSTPKDGGPGFGKAPWPGDFTTAKSSGNGAPBEERLENGDALSR 2820
Qy      2821  BSLGLPLPGSSAQPHKGLIKKCLPTISKSSLALPLLEQCTGSSRGSSASGSGGPP 2880
Db      2821  BSLGLPLPGSSAQPHKGLIKKCLPTISKSSLALPLLEQCTGSSRGSSASGSGGPP 2880
Qy      2881  RPPRQSLQOBOLNGVPIAMSIKAGTVDESSGSEFLFENFLH 2923
Db      2881  RPPRQSLQOBOLNGVPIAMSIKAGTVDESSGSEFLFENFLH 2923

```

```

RESULT 3
ABP82018
ID      ABP82018 standard; Protein; 2923 AA.
AC      ABP82018;
XX      04-MAR-2003 (first entry)
XX      Human GPCR CELSR2 protein SEQ ID NO:524.
XX

```

```

KW      G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW      G protein-coupled receptor modulator; antibody; immune-related disease;
KW      growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW      immunological-related cell proliferative disease; autoimmune disease;
KW      Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW      osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW      graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW      psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW      mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW      hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW      ulcer.
XX
OS      Homo sapiens.
XX
XX      WO200261087-A2.
XX
XX      08-AUG-2002.
XX
XX      19-DEC-2001; 2001WO-US50107.
XX
XX      19-DEC-2000; 2000US-257144P.
XX
XX      (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
XX      Burner GC, Roush CL, Brown JP;
XX
XX      WPI, 2003-046718/04.
XX      N-PSDB; AB242868.
XX
XX      New isolated antigenic peptides e.g., for G protein-coupled receptors
XX      (GPCR), useful for diagnosing and designing drugs for treating
XX      conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
XX      cancer or autoimmune diseases
XX
XX      Disclosure; Fig 1; 523pp; English.
XX
XX      The present invention describes antigenic peptides (1) comprising:
XX      (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
XX      acids. Also described: (1) an assay for the detection of a particular
XX      G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
XX      and (2) an isolated antibody having high specificity and high affinity
XX      or avidity for a particular GPCR. (1) can be used as GPCR modulators and
XX      in gene therapy. The antigenic peptides for GPCRs are useful in detecting
XX      an antibody against a particular GPCR, and in the production of specific
XX      antibodies. The peptides and antibodies are also useful for detecting the
XX      presence or absence of corresponding GPCRs. The antigenic peptides for
XX      GPCRs and antibodies are useful for diagnosing and designing drugs for
XX      treating immune-related diseases, growth-related diseases, cell
XX      regeneration-related disease, immunological-related cell proliferative
XX      diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
XX      atherosclerosis, bacterial, fungal, protozoan or viral infections,
XX      osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
XX      inflammation, allergy, Crohn's disease, diabetes, graft versus host
XX      disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
XX      anxiety, depression, schizophrenia, dementia, mental retardation, memory
XX      loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
XX      hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
XX      any other disorder in which GPCRs are involved. The antibodies may be
XX      used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode
XX      GPCR proteins given in ABP81675 to ABP82018, which are used in the
XX      exemplification of the present invention.
XX
SQ      Sequence 2923 AA;

```

```

Query Match      100.0%; Score 15545; DB 24; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MRSPATGVPPLPTPPPLLLLLLPPLLDGQVGPGRSGSGGACAPMGWLCPS 60
Db      1 MRSPATGVPPLPTPPPLLLLLLPPLLDGQVGPGRSGSGGACAPMGWLCPS 60

```

```

Qy      61 SASNMLYTRCADAGTELGHVPHHDGLRWCPSEBAHILPAPAECCPWSGRLIGIG 120

```

Db 61 SASNLMTYTSRCDAGTELTHGLVPHHDGLRVMCPSESEAH PLPBAEGCPMCSRLIGIG 120  
 QY 121 GHLSPOCKLTPREHPTCKAPRLRCOSCKLAOPGLRAGERSPESIGERRKNWNTAPQ 180  
 Db 121 GHLSPOCKLTPREHPTCKAPRLRCOSCKLAOPGLRAGERSPESIGERRKNWNTAPQ 180  
 QY 181 FQPPSYOATVPENOPACTPVASJALAIIDPDEGEAGRLTYMDALFDSRSNPFSLDPVTGA 240  
 Db 181 FQPPSYOATVPENOPACTPVASJALAIIDPDEGEAGRLTYMDALFDSRSNPFSLDPVTGA 240  
 QY 241 VTTAAEILDRCTKSTHYRVTAAODHGMRRSALATLTTLVTDTMDHPVFQOQEKSLRE 300  
 Db 241 VTTAAEILDRCTKSTHYRVTAAODHGMRRSALATLTTLVTDTMDHPVFQOQEKSLRE 300  
 QY 301 NLEGVFVLTVRATDGPAPNANILYRLGSGSGSPSEVEIDPRSGVIRTRGPVDBREV 360  
 Db 301 NLEGVFVLTVRATDGPAPNANILYRLGSGSGSPSEVEIDPRSGVIRTRGPVDBREV 360  
 QY 361 ESYQLTVEASDOGRDPGPRSTTAAVFLSVEDDNDMPQFSEKRYVQVREDVTPGAPVLK 420  
 Db 361 ESYQLTVEASDOGRDPGPRSTTAAVFLSVEDDNDMPQFSEKRYVQVREDVTPGAPVLK 420  
 QY 421 VTSASDDBKGSNAVYHSINSGNARGOFTLDAOTGALDVSPLDYETTKETTLKVRADGG 480  
 Db 421 VTSASDDBKGSNAVYHSINSGNARGOFTLDAOTGALDVSPLDYETTKETTLKVRADGG 480  
 QY 481 RPLPSNVSLGVTQVLDINDNAPIFVSTPROATVLSVPLGYTLVHVOALDADAGNARL 540  
 Db 481 RPLPSNVSLGVTQVLDINDNAPIFVSTPROATVLSVPLGYTLVHVOALDADAGNARL 540  
 QY 541 EYRLAGVGHDPPTTINNGTGMISVAELDRREVDYFVSFGVEABDHGTPALTAGASVTV 600  
 Db 541 EYRLAGVGHDPPTTINNGTGMISVAELDRREVDYFVSFGVEABDHGTPALTAGASVTV 600  
 QY 601 LDVNDNNPFTQOPEYTRLENEBAAGTSVTVSANDRNASVTTYQITSENTNRSTIS 660  
 Db 601 LDVNDNNPFTQOPEYTRLENEBAAGTSVTVSANDRNASVTTYQITSENTNRSTIS 660  
 QY 661 QSGGGLVSLALPDYKLERQYVLAVTASDGTRODTAQIVNVNTDATHRPVPOSXYTN 720  
 Db 661 QSGGGLVSLALPDYKLERQYVLAVTASDGTRODTAQIVNVNTDATHRPVPOSXYTN 720  
 QY 721 VNEBRPAGTTVLIISATDEDTGENNARTYFMEDSIFQFRIDADTGAVTTOQAELEDYDQVS 780  
 Db 721 VNEBRPAGTTVLIISATDEDTGENNARTYFMEDSIFQFRIDADTGAVTTOQAELEDYDQVS 780  
 QY 781 YTLAITARNDGIPQKSDTTYLLELVNDVNDNAPOFLRDSYQGSVYEDVPPFTSVLOISAT 840  
 Db 781 YTLAITARNDGIPQKSDTTYLLELVNDVNDNAPOFLRDSYQGSVYEDVPPFTSVLOISAT 840  
 QY 841 DRDSGLNGRVTYFQGGDDGDGDFIVESTSGIVRTLRLDRBNVAQYVLAAYVADKMP 900  
 Db 841 DRDSGLNGRVTYFQGGDDGDGDFIVESTSGIVRTLRLDRBNVAQYVLAAYVADKMP 900  
 QY 901 ARTPMEVTVTLVDVNDNPPVFEODEDFEVEENSPIGLAVARTATDPDEGTNAQIMYQI 960  
 Db 901 ARTPMEVTVTLVDVNDNPPVFEODEDFEVEENSPIGLAVARTATDPDEGTNAQIMYQI 960  
 QY 961 VEBNIEVFPOLDIFSGELTALVDLDEDRBEYVLVQATSAPVSRATYHVLRLDNDMP 1020  
 Db 961 VEBNIEVFPOLDIFSGELTALVDLDEDRBEYVLVQATSAPVSRATYHVLRLDNDMP 1020  
 QY 1021 PVLGNFEILFNNTVYNNSSFPFGAIGRVPADPDLSLTSFPERGNELSLVLNASTG 1080  
 Db 1021 PVLGNFEILFNNTVYNNSSFPFGAIGRVPADPDLSLTSFPERGNELSLVLNASTG 1080  
 QY 1081 EKLKSLALDNNRPLBAIMSVLSDGVHSTVQAQALVTYITTDMLTHSITLRLJEDMSPPR 1140  
 Db 1081 EKLKSLALDNNRPLBAIMSVLSDGVHSTVQAQALVTYITTDMLTHSITLRLJEDMSPPR 1140  
 QY 1141 FLSPLGLFIQAAVATLATPDPHVVFNVQORDTARGGHILNVSLSVGOPPGGGGPPFL 1200  
 Db 1141 FLSPLGLFIQAAVATLATPDPHVVFNVQORDTARGGHILNVSLSVGOPPGGGGPPFL 1200

Db 1141 FLSPLGLFIQAAVATLATPDPHVVFNVQORDTARGGHILNVSLSVGOPPGGGGPPFL 1200  
 QY 1201 PSEDIQERLYNRSILTLTISAORVLPDDNICTLRBCEYMMCVSLRSDSSAPFLASS 1260  
 Db 1201 PSEDIQERLYNRSILTLTISAORVLPDDNICTLRBCEYMMCVSLRSDSSAPFLASS 1260  
 QY 1261 VLFRRPIHPVGLRCHCPGFTGDYCEYEDVLCSYPCGPHGRCSREGGYTLCDGTYG 1320  
 Db 1261 VLFRRPIHPVGLRCHCPGFTGDYCEYEDVLCSYPCGPHGRCSREGGYTLCDGTYG 1320  
 QY 1321 EHCYVSAASGRCTPGVCNKGTCVNLVGGFCCDPSGDFFKPYCOVTRSPANSFTIF 1380  
 Db 1321 EHCYVSAASGRCTPGVCNKGTCVNLVGGFCCDPSGDFFKPYCOVTRSPANSFTIF 1380  
 QY 1381 RGLRORFHFTLALSTATERDGLLYNRRPNKHDPALEVLQEOVOLTFSAGESTTVS 1440  
 Db 1381 RGLRORFHFTLALSTATERDGLLYNRRPNKHDPALEVLQEOVOLTFSAGESTTVS 1440  
 QY 1441 PFPVGVSDQMHVTLQKTYNPLLGOTGLPGSPSEORVAVTVDGCDTGVALRFGSVLG 1500  
 Db 1441 PFPVGVSDQMHVTLQKTYNPLLGOTGLPGSPSEORVAVTVDGCDTGVALRFGSVLG 1500  
 QY 1501 NYSCAAQTOGGSKKSIDLTPGLLGGVPLPESPPVRRORFVGCRRMLQVDSRHIDMAD 1560  
 Db 1501 NYSCAAQTOGGSKKSIDLTPGLLGGVPLPESPPVRRORFVGCRRMLQVDSRHIDMAD 1560  
 QY 1561 FIANNGTVPCCAKKNVCDNSNTHNGGTICVNMOMDAPCECPJGFGGKSCAQMAMPQHL 1620  
 Db 1561 FIANNGTVPCCAKKNVCDNSNTHNGGTICVNMOMDAPCECPJGFGGKSCAQMAMPQHL 1620  
 QY 1621 GSSLVAMHGLSLPISQPMYLSIMFRTROADVLLQAITRGRSTITLQLEBGMVLSVBGT 1680  
 Db 1621 GSSLVAMHGLSLPISQPMYLSIMFRTROADVLLQAITRGRSTITLQLEBGMVLSVBGT 1680  
 QY 1681 GLQASSLLEBRANDGEMHHAQLALGASGGGHAHLSFDGQQAENLGRHLGHLIS 1740  
 Db 1681 GLQASSLLEBRANDGEMHHAQLALGASGGGHAHLSFDGQQAENLGRHLGHLIS 1740  
 QY 1741 NITVGGIGPGAGVARGRGCLQGVAVSDTPEGVNSLDPSSHESINVEQGSCLPDCSN 1800  
 Db 1741 NITVGGIGPGAGVARGRGCLQGVAVSDTPEGVNSLDPSSHESINVEQGSCLPDCSN 1800  
 QY 1801 PCPANSYCSNMDYSYSCDPEGYGDNCTNVCNLPCEHOSYCTRKP9APHGTCPCPN 1860  
 Db 1801 PCPANSYCSNMDYSYSCDPEGYGDNCTNVCNLPCEHOSYCTRKP9APHGTCPCPN 1860  
 QY 1861 YLGPYCETRLIDPCCRGMMGHPTGCPNCNDVSKGDPDCNKTSGRCHCKENHYRPGSPT 1920  
 Db 1861 YLGPYCETRLIDPCCRGMMGHPTGCPNCNDVSKGDPDCNKTSGRCHCKENHYRPGSPT 1920  
 QY 1921 CLLCDYPTGSLSRVCDPBDQCPCPKPGVIGRQCDRCNPNFAEVYTNNGCEVYDSCPRAI 1980  
 Db 1921 CLLCDYPTGSLSRVCDPBDQCPCPKPGVIGRQCDRCNPNFAEVYTNNGCEVYDSCPRAI 1980  
 QY 1981 EAGIWWPRTTRGLPAAPCPYKGSFGTAVRHCDBERGMPLPMLFNCTSTTFSELKCPABRL 2040  
 Db 1981 EAGIWWPRTTRGLPAAPCPYKGSFGTAVRHCDBERGMPLPMLFNCTSTTFSELKCPABRL 2040  
 QY 2041 QNBSGLSGRSQOLALLNNAHTAGYPGSDVVAVALQTLRLAHSTORGFLSATO 2100  
 Db 2041 QNBSGLSGRSQOLALLNNAHTAGYPGSDVVAVALQTLRLAHSTORGFLSATO 2100  
 QY 2101 DVHFTENLLRVGSALLDTRANKRMELIOTBTGTMALLQHYEAVASALANONRHITYLSPF 2160  
 Db 2101 DVHFTENLLRVGSALLDTRANKRMELIOTBTGTMALLQHYEAVASALANONRHITYLSPF 2160  
 QY 2161 TTVTNTVIVSVRLDKGNPAKULPRYALRGEOPDLLETVIILBSVFRBTPVVRPAG 2220  
 Db 2161 TTVTNTVIVSVRLDKGNPAKULPRYALRGEOPDLLETVIILBSVFRBTPVVRPAG 2220  
 QY 2221 PGEAQPEELARORRHELSCGEAVASVITYRTLAGLLPHNYDDPKSLRVKXPIINT 2280  
 Db 2221 PGEAQPEELARORRHELSCGEAVASVITYRTLAGLLPHNYDDPKSLRVKXPIINT 2280

```

QY 2281 PVSISVHDEDELLPRLADKPVTVQRLLETERTTPICVFNHNSILVSGTGMARGCE 2340
DB 2281 PVSISVHDEDELLPRLADKPVTVQRLLETERTTPICVFNHNSILVSGTGMARGCE 2340
QY 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLVYVALGVTLAALLTFFPFLU 2400
DB 2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPLKTLVYVALGVTLAALLTFFPFLU 2400
QY 2401 LRLRNSNGHGRIRNLTALGLAQVLVLGINQADLPFACTVAILLHFLYLCFSPWALIE 2460
DB 2401 LRLRNSNGHGRIRNLTALGLAQVLVLGINQADLPFACTVAILLHFLYLCFSPWALIE 2460
QY 2461 ALHLVYALTEVADVNTGPMRFYTMCGWCPAPFTTGLAVGLDPGYCNPFCWLSTYDTLI 2520
DB 2461 ALHLVYALTEVADVNTGPMRFYTMCGWCPAPFTTGLAVGLDPGYCNPFCWLSTYDTLI 2520
QY 2521 WSPFAGVAVAVSMVFLYTLAARASCAORQGEKKGPVSGLOPSPAVILLLSATWLLAL 2580
DB 2521 WSPFAGVAVAVSMVFLYTLAARASCAORQGEKKGPVSGLOPSPAVILLLSATWLLAL 2580
QY 2581 LSVNSDTLLFHYLFATCNCIQGPFIPLSYVVLSEVKRKLKLACSRKPSPPDPAITTKSTL 2640
DB 2581 LSVNSDTLLFHYLFATCNCIQGPFIPLSYVVLSEVKRKLKLACSRKPSPPDPAITTKSTL 2640
QY 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTSRGSKSQPSYIPLAREESALNPGQGPGLG 2700
DB 2641 TSSYNCPSPYADGRLYQPYGDSAGSLHSTSRGSKSQPSYIPLAREESALNPGQGPGLG 2700
QY 2701 DPGSLFLBQDOOHDPDTSDDLSELDPOSGSYASTHSDSEEEEEEEREAAPGREG 2760
DB 2701 DPGSLFLBQDOOHDPDTSDDLSELDPOSGSYASTHSDSEEEEEEEREAAPGREG 2760
QY 2761 WDSLLGPGAERLPLHSTPKDGGPGPKAPWPGDFTTAKESSGNGAPBEERLRENGDALSR 2820
DB 2761 WDSLLGPGAERLPLHSTPKDGGPGPKAPWPGDFTTAKESSGNGAPBEERLRENGDALSR 2820
QY 2821 ESSLGPELPSSSAQPHNGILLKCKLPITSEKSSILRLPLEQCTGSSRGSSASGSRGGPPP 2880
DB 2821 ESSLGPELPSSSAQPHNGILLKCKLPITSEKSSILRLPLEQCTGSSRGSSASGSRGGPPP 2880
QY 2881 RPPRQSLQEQNLGWPVIMSIKAGTVDESSGSEPLFPNPLH 2923
DB 2881 RPPRQSLQEQNLGWPVIMSIKAGTVDESSGSEPLFPNPLH 2923

```

RESULT 4  
AAU07053  
ID AAU07053 standard; Protein; 2956 AA.

24-OCT-2001 (fixed entry)

Human Flamingo polypeptide.

Flamingo; human; G-protein coupled receptor; diabetes; protozoan; cancer; signal transduction pathway; bacterial; fungal; viral; anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV; obesity; anxiety; hypertension; hypertension; urinary retention; angina pectoris; allergy; myocardial infarction; stroke; ulcer; schizophrenia; psychotic disorder; benign prostatic hypertrophy; neurological disorder; manic depression; delirium; dementia; severe mental retardation; Huntington's disease; Gilles de la Tourette's syndrome; antibacterial; antiviral; antiprotocozal; anti-HIV; anorectic; antiasthmatic; antiparkinsonian; cardiac; cerebroprotective; neuroprotective; antidepressant; anticonvulsant; antisense therapy; gene therapy.

OS Homo sapiens.  
XX WO200161003-A1.  
XX 23-AUG-2001.  
XX PD

```

XX 19-FEB-2001; 2001WO-GB00680.
PF 19-FEB-2001; 2000GB-0004196.
PR 19-FEB-2001; 2000GB-0004196.
XX (SMK ) SMITHKLINE BEECHAM PLC.
PA Testa TT;
PI WPI; 2001-502792/55.
XX N-PSDB; AAS11677.
DR An isolated Flamingo polypeptide useful for treating diseases such as
PT HIV, cancer, asthma, Parkinson's disease, acute heart failure,
PT osteoporosis -
PS Claim 2; Page 28-29; 66pp; English.
XX The sequence represents a human Flamingo polypeptide. Flamingo is a
CC member of the G-protein coupled receptor family, which is involved in
CC signal transduction pathways. By screening to identify compounds that
CC stimulate or inhibit the function or level of the protein, treatments can
CC be developed for various diseases and bacterial, fungal, protozoan and
CC viral infections, including HIV, cancer, diabetes, obesity, anorexia,
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypertension,
CC stroke, ulcers, allergies and benign prostatic hypertrophy. Also
CC treatable are psychotic and neurological disorders such as anxiety,
CC schizophrenia, manic depression, delirium, dementia, severe mental
CC retardation, Huntington's disease and Gilles de la Tourette's syndrome.
XX Sequence 2956 AA;

```

Query Match 99.8%; Score 1518.5; DB 22; Length 2956;  
Best Local Similarity 98.9%; Pred. No. 0; Indels 33; Gaps 1;  
Matches 2923; Conservative 0; Mismatches 0;

```

QY 1 MRSPATGVPLPTPPPLLLLLLLPPLLDGVPCSLGSRGSSGACAPMGWLCPS 60
DB 1 MRSPATGVPLPTPPPLLLLLLLPPLLDGVPCSLGSRGSSGACAPMGWLCPS 60
QY 61 SASNMLYTSRCRAGTTLTHLVPHHDGLRWCPBSAAHPLPAPPGCWCSCLLGIG 120
DB 61 SASNMLYTSRCRAGTTLTHLVPHHDGLRWCPBSAAHPLPAPPGCWCSCLLGIG 120
QY 61 SASNMLYTSRCRAGTTLTHLVPHHDGLRWCPBSAAHPLPAPPGCWCSCLLGIG 120
DB 61 SASNMLYTSRCRAGTTLTHLVPHHDGLRWCPBSAAHPLPAPPGCWCSCLLGIG 120
QY 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRGERSPEBSLGRRKRNVTAPQ 180
DB 121 GHLSPOGKLTPEEHPCLKAPRLRCQSCKLAQAGLRGERSPEBSLGRRKRNVTAPQ 180
QY 181 FQPSYQATVBNOPAGTPVVALRAIDPDEAGRLLEYMDALFDSRSNQPFSIDPVTGA 240
DB 181 FQPSYQATVBNOPAGTPVVALRAIDPDEAGRLLEYMDALFDSRSNQPFSIDPVTGA 240
QY 241 VTTABEIDRETKSTHVVRVTAQDHGMPRRSALATLTLLVTNDHDPFEOEYKESIRE 300
DB 241 VTTABEIDRETKSTHVVRVTAQDHGMPRRSALATLTLLVTNDHDPFEOEYKESIRE 300
QY 301 NLEFGEVFLYTRATDGAAPPAANILYRLBSGSGSPSVFPIIDRSQYIRRGVNDREEV 360
DB 301 NLEFGEVFLYTRATDGAAPPAANILYRLBSGSGSPSVFPIIDRSQYIRRGVNDREEV 360
QY 361 ESYQUTVEASDQDGPGRSTTAAVFLSVBDNDNABQFSEKRYVVOVREDVTPGAPYLR 420
DB 361 ESYQUTVEASDQDGPGRSTTAAVFLSVBDNDNABQFSEKRYVVOVREDVTPGAPYLR 420
QY 421 VTAADPRKGSNAVHYHISMGNNAGQYLLAQGTALVVSPLDYETTKYTLRRADQGG 480
DB 421 VTAADPRKGSNAVHYHISMGNNAGQYLLAQGTALVVSPLDYETTKYTLRRADQGG 480
QY 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLSVPLGYLVLAHQADADAGDNARL 540
DB 481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLSVPLGYLVLAHQADADAGDNARL 540

```

QY 541 ERLAGVGHDPPTINNGTGMISVAELREBVDYSPGVEARDHGTPTALITASASVTV 600  
 DB 541 ERLAGVGHDPPTINNGTGMISVAELREBVDYSPGVEARDHGTPTALITASASVTV 600  
 QY 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSVTVYITISGNTNRNPSITS 660  
 DB 601 LDVNDNNPFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSVTVYITISGNTNRNPSITS 660  
 QY 661 QSGGGLVSLALPDYTYLEROYVLAATASDPTRODTAQIVNVTVDATHTHPVFOSSHXTN 720  
 DB 661 QSGGGLVSLALPDYTYLEROYVLAATASDPTRODTAQIVNVTVDATHTHPVFOSSHXTN 720  
 QY 721 VNEDRAGTTVVILSATDEDTGENARITYPMEDSIPQFRIDADTGAVTTOABLDYEDYS 780  
 DB 721 VNEDRAGTTVVILSATDEDTGENARITYPMEDSIPQFRIDADTGAVTTOABLDYEDYS 780  
 QY 781 YTLATARNGIPQKSDTTYLETUNVDNDNAPQFLRDSYQSVYEDVPPFTSVLQISAT 840  
 DB 781 YTLATARNGIPQKSDTTYLETUNVDNDNAPQFLRDSYQSVYEDVPPFTSVLQISAT 840  
 QY 841 DRDSGLNGRVFTYFQGGDDGDGPIVESTSGIVRTLRRLDRNVQAQYVLAAYDKMP 900  
 DB 841 DRDSGLNGRVFTYFQGGDDGDGPIVESTSGIVRTLRRLDRNVQAQYVLAAYDKMP 900  
 QY 901 ARTMEVTVTVLDVNDNPPVFEQDEPVEBENSPIGLAVARVATDPDEGNAQIMYQI 960  
 DB 901 ARTMEVTVTVLDVNDNPPVFEQDEPVEBENSPIGLAVARVATDPDEGNAQIMYQI 960  
 QY 961 VEGNPEVFOLODFSGELTALVDYEDBPEYLVYQATSAPVSAATVHNLDRNDP 1020  
 DB 961 VEGNPEVFOLODFSGELTALVDYEDBPEYLVYQATSAPVSAATVHNLDRNDP 1020  
 QY 1021 PVLNGEILFNNVYTRSSSPFGAIGRPVADPDIDSLITYSFERGNEILVTLNASTG 1080  
 DB 1021 PVLNGEILFNNVYTRSSSPFGAIGRPVADPDIDSLITYSFERGNEILVTLNASTG 1080  
 QY 1081 ELKLSALDNNRPLEAIVSVSDGVHSTVTAOCALRTIITDEMTLHSTITLIEDMSPER 1140  
 DB 1081 ELKLSALDNNRPLEAIVSVSDGVHSTVTAOCALRTIITDEMTLHSTITLIEDMSPER 1140  
 QY 1141 FLSPLGLFLOAVATLAPPHDVVFNVOPTDAGGHLNVLSTVGPFGGSGPPL 1200  
 DB 1141 FLSPLGLFLOAVATLAPPHDVVFNVOPTDAGGHLNVLSTVGPFGGSGPPL 1200  
 QY 1201 PSEDLORLYLNSLTALISAGRVLPDDNICLARECENYMRCSVTLRDSAPFIASSS 1260  
 DB 1201 PSEDLORLYLNSLTALISAGRVLPDDNICLARECENYMRCSVTLRDSAPFIASSS 1260  
 QY 1261 VLFRLPHVVGGLRCRCPGFTGDCYCTEVDLCYSRPGPHGRCSRREGGYTCLCRDGYTG 1320  
 DB 1261 VLFRLPHVVGGLRCRCPGFTGDCYCTEVDLCYSRPGPHGRCSRREGGYTCLCRDGYTG 1320  
 QY 1321 EHCEVARSGRCTPGVCKNGGTGMNLLVGGFKDCSGSPEKXYCOVTRSPRHSFITF 1380  
 DB 1321 EHCEVARSGRCTPGVCKNGGTGMNLLVGGFKDCSGSPEKXYCOVTRSPRHSFITF 1380  
 QY 1381 RGLRORFHTLALSPATKERDGLLNGRFXEKHDFVLALEVOBQVOLFPSAGESITTVS 1440  
 DB 1381 RGLRORFHTLALSPATKERDGLLNGRFXEKHDFVLALEVOBQVOLFPSAGESITTVS 1440  
 QY 1441 PFLPGVSDGQMTTVOLKYNKPLLGQTCIPGSPSEKVAVTVTDGCTGVALRFGSVTG 1500  
 DB 1441 PFLPGVSDGQMTTVOLKYNKPLLGQTCIPGSPSEKVAVTVTDGCTGVALRFGSVTG 1500  
 QY 1501 NYSCAAGTQGGSKSLDLTGPLLLGGVPLPESFPBAMQFVGCNRNTQVDSRHIDMD 1560  
 DB 1501 NYSCAAGTQGGSKSLDLTGPLLLGGVPLPESFPBAMQFVGCNRNTQVDSRHIDMD 1560  
 QY 1561 FIANNGTVPCCPAKXVCDSNTCHNGTCVNWDAFSCCEPLGFGKSCAQEMANPOHFL 1620  
 DB 1561 FIANNGTVPCCPAKXVCDSNTCHNGTCVNWDAFSCCEPLGFGKSCAQEMANPOHFL 1620  
 QY 1621 GSSLVAMHGLSLPISQWYLSLMFRTROADGVLLQAITRGRSTITLQABEGHVLVSGT 1680

DB 1621 GSSLVAMHGLSLPISQWYLSLMFRTROADGVLLQAITRGRSTITLQABEGHVLVSGT 1680  
 QY 1681 GLQASSLLEBRANDGMHHAQALAGSGPGRHILSPDYGOQRAEENLPRHGLHLS 1740  
 DB 1681 GLQASSLLEBRANDGMHHAQALAGSGPGRHILSPDYGOQRAEENLPRHGLHLS 1740  
 QY 1741 NITVGGIPGAPGVARGRGLQGVASVDTPEGVNSLDPHSGESINVEQCSLPPCCSN 1800  
 DB 1741 NITVGGIPGAPGVARGRGLQGVASVDTPEGVNSLDPHSGESINVEQCSLPPCCSN 1800  
 QY 1801 PCPANSYCSNDWDSYSCDPEYTGDNCTNVCDLNPCEHQSCTRKPSAPHGYTCECPN 1860  
 DB 1801 PCPANSYCSNDWDSYSCDPEYTGDNCTNVCDLNPCEHQSCTRKPSAPHGYTCECPN 1860  
 QY 1861 YLGPYCEFRIDOPCRGMWGHPTGCPKCDVSKGDPUCNKTSGCHCKENHYRPGSP 1920  
 DB 1861 YLGPYCEFRIDOPCRGMWGHPTGCPKCDVSKGDPUCNKTSGCHCKENHYRPGSP 1920  
 QY 1921 CLICDCTYPTGSLSRVCDPEBQCECKPQVIGRQCDRCNDPFAEYTTNGCEVNYDSCPPAI 1980  
 DB 1921 CLICDCTYPTGSLSRVCDPEBQCECKPQVIGRQCDRCNDPFAEYTTNGCEVNYDSCPPAI 1980  
 QY 1981 EAGIWPPTRRFGLPAAAPCPFGSGTAVRHCDBERGMLPMLFNCTSTTSEELKGAERL 2040  
 DB 1981 EAGIWPPTRRFGLPAAAPCPFGSGTAVRHCDBERGMLPMLFNCTSTTSEELKGAERL 2040  
 QY 2041 QNBSGLDSGRSQQALALLRNAHTAGYFGSDVKAVALQTLRLAHSTGRGELSTQ 2100  
 DB 2041 QNBSGLDSGRSQQALALLRNAHTAGYFGSDVKAVALQTLRLAHSTGRGELSTQ 2100  
 QY 2101 DVHFTENLARGSALDITANKRHEMLIQTEGCTAMLQHYEAVALAQNMRHTVLSPF 2160  
 DB 2101 DVHFTENLARGSALDITANKRHEMLIQTEGCTAMLQHYEAVALAQNMRHTVLSPF 2160  
 QY 2161 TIVPNTVITSVRLDKGNPAKAPRYBALRGEOPDLETTVILPESIFRETTPVVRPAG 2220  
 DB 2161 TIVPNTVITSVRLDKGNPAKAPRYBALRGEOPDLETTVILPESIFRETTPVVRPAG 2220  
 QY 2221 PGBAOPREELARORRHELSQGEAVASVITYRTLAGLLPHNYOPDKSLVPRKPIINT 2280  
 DB 2221 PGBAOPREELARORRHELSQGEAVASVITYRTLAGLLPHNYOPDKSLVPRKPIINT 2280  
 QY 2281 PVSISVHDBEELLPRALDKPEVTVOFLLSTBERTKPICVFNHNSILVSGTGSARCE 2340  
 DB 2281 PVSISVHDBEELLPRALDKPEVTVOFLLSTBERTKPICVFNHNSILVSGTGSARCE 2340  
 QY 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGEILPKLTLYVALGVTLAALLTFEFLT 2400  
 DB 2341 VVFRNESHVSQCNHMTSPATLMDVSRRENGEILPKLTLYVALGVTLAALLTFEFLT 2400  
 QY 2401 LRIILSNHGIRBNLTALAGLAQVFLGIGNOADIPLACTVIALILHFLYCTFSMALLE 2460  
 DB 2401 LRIILSNHGIRBNLTALAGLAQVFLGIGNOADIPLACTVIALILHFLYCTFSMALLE 2460  
 QY 2461 ALHLRYALTBRVDVNTGEMRFYMLGMGVPAFITGLAVGLDPEGNDPFCMLSTYDTLI 2520  
 DB 2461 ALHLRYALTBRVDVNTGEMRFYMLGMGVPAFITGLAVGLDPEGNDPFCMLSTYDTLI 2520  
 QY 2521 WSPAPVPAFAYMSVFLYTLAARASCAAOROGFEKKGVSGLQBSFALLLSATWLLAL 2580  
 DB 2521 WSPAPVPAFAYMSVFLYTLAARASCAAOROGFEKKGVSGLQBSFALLLSATWLLAL 2580  
 QY 2581 LSVNSDTLLFHTLFAATCNCIOGPFIFLSYVVLSEKVRALKLACSRKSPDPALTTKSTL 2640  
 DB 2581 LSVNSDTLLFHTLFAATCNCIOGPFIFLSYVVLSEKVRALKLACSRKSPDPALTTKSTL 2640  
 QY 2641 TSSYNCPSPYADGRLYQYGYDSAGSLHSTSRGSKSQPSYIFPLLRBEBALMPGOCPGIG 2700  
 DB 2641 TSSYNCPSPYADGRLYQYGYDSAGSLHSTSRGSKSQPSYIFPLLRBEBALMPGOCPGIG 2700  
 QY 2701 DPGSLFLGGOQOQHDPDUDSDSLBDDOSGSVASSTSSDSEBEEBEEBAAAPRGOG 2760

Db 2701 DRGSLFLEGGDQHDPTDSDSLSDLDQSGSYASTHSDSEBEEBEEBAPPGG 2760  
Qy 2761 WDLILGGAERILPLHSTPDGPGPGKAWPDGFTTAKSSGNGAPBERRLENGDALSR 2820  
Db 2761 WDLILGGAERILPLHSTPDGPGPGKAWPDGFTTAKSSGNGAPBERRLENGDALSR 2820  
Qy 2821 BSLGLPLPSSSQPKH-----GILKKCLPTI 2847  
Db 2821 BSLGLPLPSSSQPKHGWGTPTSCRAVLVSLIPIILMPLHTAPPRPTGIKKCLPTI 2880  
Qy 2848 SEKSSILRLPLEOCTGSSRGSSASBEGSRGPPPPRPSLOEOLNGVPMIAMSIRAGTV 2907  
Db 2881 SEKSSILRLPLEOCTGSSRGSSASBEGSRGPPPPRPSLOEOLNGVPMIAMSIRAGTV 2940  
Qy 2908 DDDSGSEPLFFNFYH 2923  
Db 2941 DDDSGSEPLFFNFYH 2956

RESULT 5  
ABU11556  
ID ABU11556 standard; Protein; 2894 AA.  
AC ABU11556;  
XX  
XX  
DT 12-FEB-2003 (first entry)  
DE Human MDDT polypeptide SEQ ID 503.  
XX  
XX  
MDDT; human; disease detection and treatment molecule polypeptide;  
anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
haemostatic; nephrotropic; anti-nausea; anti-psoriasis; hepatotropic;  
gene therapy; protein replacement therapy; cell proliferative disorder;  
cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;  
psoriasis; hepatitis.  
XX  
XX  
Homo sapiens.  
XX  
XX  
FN W0200279449-A2.  
XX  
XX  
PD 10-OCT-2002.  
XX  
XX  
PE 27-MAR-2002; 2002MO-US09944.  
XX  
XX  
PR 28-MAR-2001; 2001US-279619P.  
PR 29-MAR-2001; 2001US-280067P.  
PR 29-MAR-2001; 2001US-280068P.  
PR 16-MAY-2001; 2001US-291280P.  
PR 17-MAY-2001; 2001US-291289P.  
PR 17-MAY-2001; 2001US-291849P.  
PR 19-JUN-2001; 2001US-299428P.  
PR 20-JUN-2001; 2001US-299776P.  
PR 20-JUN-2001; 2001US-300001P.  
XX  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX  
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J.  
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;  
PI Daugherty SC, Dam TC, Liu TP, Nguyen DA, Kleeefeld Y, Gerstein EH;  
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
PI Flores V, Marwala R, Lo A, Lan RY, Urashka ME;  
XX  
XX  
DR MPI; 2003-058431/05.  
XX  
XX  
DR N-PEDB; ABX34546.  
XX  
XX  
PT New purified disease detection and treatment molecule proteins and  
PT polynucleotides, useful for diagnosing, treating or preventing cancers  
PT (e.g. leukaemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis  
PT or hepatitis  
XX  
XX  
PS Claim 27; SEQ ID NO 503; 339pp + Sequence Listing; English.

XX This invention describes a novel disease detection and treatment molecule  
CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
CC anti-nausea, anti-psoriasis and hepatotropic activity. The polynucleotides  
CC and the polypeptides of the invention can be used for gene therapy,  
CC protein replacement therapy and are useful for treating a variety of  
CC diseases or conditions. These polypeptides or polynucleotides are  
CC particularly useful for diagnosing, treating or preventing cell  
CC proliferative disorders (e.g. cancers including adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded  
CC by ABU11450-ABU11845, described in the disclosure of the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Sequence 2894 AA;

Query Match 99.0%; Score 15393; DB 24; Length 2894;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 LGDQVGPCRSIGSRGSSGACAPMGWLCPSSASNTMLYTSRCRDAGTELTHGLVPHHDG 89  
Db 1 LGDQVGPCRSIGSRGSSGACAPMGWLCPSSASNTMLYTSRCRDAGTELTHGLVPHHDG 60  
Qy 90 LRWVCPSEBAIPIPPAPGEGPCRLIGTGHISPOGKTLTPBHPCLKAPRRCOSCK 149  
Db 61 LRWVCPSEBAIPIPPAPGEGPCRLIGTGHISPOGKTLTPBHPCLKAPRRCOSCK 120  
Qy 150 LAQAPGLRAGRSSEBSIGRRKKNVNTAPQFOPSPYQATVBNOPAGTPASIRAIQDP 209  
Db 121 LAQAPGLRAGRSSEBSIGRRKKNVNTAPQFOPSPYQATVBNOPAGTPASIRAIQDP 180  
Qy 210 EGBAGRLTYMDALFDRSNQPFSLDPVTGAVTAAEILDRKTSYHVRVTAQDHGERR 269  
Db 181 EGBAGRLTYMDALFDRSNQPFSLDPVTGAVTAAEILDRKTSYHVRVTAQDHGERR 240  
Qy 270 SALATLTILVTPTDHPDVPVPOQYKESLRNLEVGVEYTLVRAITDGPANNAITRL 329  
Db 241 SALATLTILVTPTDHPDVPVPOQYKESLRNLEVGVEYTLVRAITDGPANNAITRL 300  
Qy 330 EGSQSPSEVFEIDPRSGVIRTRGPVDRBEVESYQLTVASDQGRDPGPRSTAAVFLSV 389  
Db 301 EGSQSPSEVFEIDPRSGVIRTRGPVDRBEVESYQLTVASDQGRDPGPRSTAAVFLSV 360  
Qy 390 EDDNDNAPQSEKRYVQVREDDVTGAPVLVTASDRDKGSNAVHYHSIMSNGARQFYL 449  
Db 361 EDDNDNAPQSEKRYVQVREDDVTGAPVLVTASDRDKGSNAVHYHSIMSNGARQFYL 420  
Qy 450 DAQFGALDVSPIDYETTKKTYTLRVAQDQGRPPLSVSGLYTVQVLDINDNAITPSTP 509  
Db 421 DAQFGALDVSPIDYETTKKTYTLRVAQDQGRPPLSVSGLYTVQVLDINDNAITPSTP 480  
Qy 510 FOATVLESVPAGYVLAFOAIDADAGNARLEVLAGVGDPPPTINGGWLIVAAELD 569  
Db 481 FOATVLESVPAGYVLAFOAIDADAGNARLEVLAGVGDPPPTINGGWLIVAAELD 540  
Qy 570 REEVDFTSPGVARADHGTPLATASASVTVLDVNDNPTFTQBEYTVRLNEDAAVQTSV 629  
Db 541 REEVDFTSPGVARADHGTPLATASASVTVLDVNDNPTFTQBEYTVRLNEDAAVQTSV 600  
Qy 630 VTYSADVADHASYTVQITSGNTRNRPSTISQSGGIVSALPLDYLERQYVLAATASD 689  
Db 601 VTYSADVADHASYTVQITSGNTRNRPSTISQSGGIVSALPLDYLERQYVLAATASD 660  
Qy 690 GTRQDTQIIVVNTDANTHRPVPQSHYTVNVEDRAGTGVVILISATDDEGENARITY 749  
Db 661 GTRQDTQIIVVNTDANTHRPVPQSHYTVNVEDRAGTGVVILISATDDEGENARITY 720



QY 750 FMEBISIPQFIIDADTGAVTTOAEIDYEDQVSYTLATARDNGIPQKSDTTYLBIVNDV 809  
DB 721 FMEBISIPQFIIDADTGAVTTOAEIDYEDQVSYTLATARDNGIPQKSDTTYLBIVNDV 780  
QY 810 DNAPQFRLDSYQSGSVEDVPEFTSVLOISATDRSGNGRVEFTYFOGDDGDDDFIYEST 869  
DB 781 DNAPQFRLDSYQSGSVEDVPEFTSVLOISATDRSGNGRVEFTYFOGDDGDDDFIYEST 840  
QY 870 SGIVTRLRLRLDRBNAQVYLRAVAVDKGMPARTPMEVTVTVLVNDNPPVFEQDEDFV 929  
DB 841 SGIVTRLRLRLDRBNAQVYLRAVAVDKGMPARTPMEVTVTVLVNDNPPVFEQDEDFV 900  
QY 930 VEERNSPGLAVATYATDPBGTNAQIMYOI VEGNIPVFOLOI FSGELTALVDLYEDR 989  
DB 901 VEERNSPGLAVATYATDPBGTNAQIMYOI VEGNIPVFOLOI FSGELTALVDLYEDR 960  
QY 990 PEYVLVQATSAPLVSAATVHVRLDRNDNPVYLGNEILLFNNYVNTRSSPFGAIGRY 1049  
DB 961 PEYVLVQATSAPLVSAATVHVRLDRNDNPVYLGNEILLFNNYVNTRSSPFGAIGRY 1020  
QY 1050 PAHDPOISDSLYTSFERGNEISLVTLNASTGBLKLSBALDNNRPLBAIMSVLVSDGVHAY 1109  
DB 1021 PAHDPOISDSLYTSFERGNEISLVTLNASTGBLKLSBALDNNRPLBAIMSVLVSDGVHAY 1080  
QY 1110 TAOCALAVTITTDMLTHSITTLRLMEDSPEFTSLPLGLFTIQAATAATLAPPDHVVENV 1169  
DB 1081 TAOCALAVTITTDMLTHSITTLRLMEDSPEFTSLPLGLFTIQAATAATLAPPDHVVENV 1140  
QY 1170 QRTDAPAGHILNVLNLSVGPQPGGPPFLPSEDLQERLYLNRSLTATSAORVLPEDD 1229  
DB 1141 QRTDAPAGHILNVLNLSVGPQPGGPPFLPSEDLQERLYLNRSLTATSAORVLPEDD 1200  
QY 1230 NICLRBCEVNYMRCVSVLAFDSSAPFIASSSVLERPIHPVGLRCRCRCPGTGDCYETEV 1289  
DB 1201 NICLRBCEVNYMRCVSVLAFDSSAPFIASSSVLERPIHPVGLRCRCRCPGTGDCYETEV 1260  
QY 1290 DLYCSRRCGHHGCRSREGGTYCLCRDGYTGENHCVASRGCTPRGCKNGCTVNLVNG 1349  
DB 1261 DLYCSRRCGHHGCRSREGGTYCLCRDGYTGENHCVASRGCTPRGCKNGCTVNLVNG 1320  
QY 1350 GFKDCDPSGDFEKPCVQVTTSPFAHSPIIFRGILRQGFHFLALSFYTKERDGLLNGR 1409  
DB 1321 GFKDCDPSGDFEKPCVQVTTSPFAHSPIIFRGILRQGFHFLALSFYTKERDGLLNGR 1380  
QY 1410 FNEKHDFALEVOIYQVOLTFSAGESTTVSPVPGSVSDGQWHTVOLKTYNRPILGQTS 1469  
DB 1381 FNEKHDFALEVOIYQVOLTFSAGESTTVSPVPGSVSDGQWHTVOLKTYNRPILGQTS 1440  
QY 1470 LPOGSPBOKAVVTVVDCDGVALRFGSVLGNTSCAAQGTQGSKSLDTGPIILGQVP 1529  
DB 1441 LPOGSPBOKAVVTVVDCDGVALRFGSVLGNTSCAAQGTQGSKSLDTGPIILGQVP 1500  
QY 1530 DLBESFVNRAROFVGCARNIQVDSRHIDMADFIANNTPGCPAKKNVCSNTCHNGGTC 1589  
DB 1501 DLBESFVNRAROFVGCARNIQVDSRHIDMADFIANNTPGCPAKKNVCSNTCHNGGTC 1560  
QY 1590 VNOMDAPSCCPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPMYLSIMFTRQA 1649  
DB 1561 VNOMDAPSCCPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPMYLSIMFTRQA 1620  
QY 1650 DGVLLQAITGRSTITLQLRBGHVMLSVBGTGLOASLSRLBPGANDGMHQAOLALGAS 1709  
DB 1621 DGVLLQAITGRSTITLQLRBGHVMLSVBGTGLOASLSRLBPGANDGMHQAOLALGAS 1680  
QY 1710 GGPCHAILLSFDYGGQARAGNLPRLHGLHLSNITVGGIPGPAGVANGFGGCLQGYRVSD 1769  
DB 1681 GGPCHAILLSFDYGGQARAGNLPRLHGLHLSNITVGGIPGPAGVANGFGGCLQGYRVSD 1740  
QY 1770 TPBENVNLDSHGSINVEOGCSLPDPCDNPCCANYSICNDMSYSCSDPGYTGNCNT 1829  
DB 1741 TPBENVNLDSHGSINVEOGCSLPDPCDNPCCANYSICNDMSYSCSDPGYTGNCNT 1800  
QY 1830 NVCDLNPCEHQSVCTKPSAPHGVTCEPNTYLGPCETRIDQCPRGWGHPTGCPNC 1889

DB 1801 NVCDLNPCEHQSVCTKPSAPHGVTCEPNTYLGPCETRIDQCPRGWGHPTGCPNC 1860  
QY 1890 DYSKGDPPCNTSGCHCKENHVRPGSPFTCLCDQCVTGSLSRVCDEDDQCPCKPBY 1949  
DB 1861 DYSKGDPPCNTSGCHCKENHVRPGSPFTCLCDQCVTGSLSRVCDEDDQCPCKPBY 1920  
QY 1950 IGRQCDRCNPPAEVTVNGCEVNYDSCPRALBAGIWMPPTRGFLPAAAPCPKSPFGTAVR 2009  
DB 1921 IGRQCDRCNPPAEVTVNGCEVNYDSCPRALBAGIWMPPTRGFLPAAAPCPKSPFGTAVR 1980  
QY 2010 HCDHERGMLPENVFNCTSTTSBLKGAERLORNBSSGLDSGSQOLALLRNATQHTAGY 2069  
DB 1981 HCDHERGMLPENVFNCTSTTSBLKGAERLORNBSSGLDSGSQOLALLRNATQHTAGY 2040  
QY 2070 FGSQDVAVAYQATRLLAHSTQRGGLSATQDVHFENMLARGSALLDPANRHMELIQO 2129  
DB 2041 FGSQDVAVAYQATRLLAHSTQRGGLSATQDVHFENMLARGSALLDPANRHMELIQO 2100  
QY 2130 TBEGTAMLLQHYEAVASALAQNMRRHTYLPFTIVTPNIVISVVRLDKGNFAGAKLPRYEA 2189  
DB 2101 TBEGTAMLLQHYEAVASALAQNMRRHTYLPFTIVTPNIVISVVRLDKGNFAGAKLPRYEA 2160  
QY 2190 LAGEOPPDLETTYVILPESVFRSTPPVVRPAGGGAQOEPEBLARQRHPELSQGEAVASV 2249  
DB 2161 LAGEOPPDLETTYVILPESVFRSTPPVVRPAGGGAQOEPEBLARQRHPELSQGEAVASV 2220  
QY 2250 IYRTLAGLLPHNYDDKSLRVPRKPIINTPVASISVHDEBELPRALDKVPTVOFRLL 2309  
DB 2221 IYRTLAGLLPHNYDDKSLRVPRKPIINTPVASISVHDEBELPRALDKVPTVOFRLL 2280  
QY 2310 ETEERTKPICVFNWNSILVSGTGSARGCEVVFNRSHVSQCNHMTSPAVLMDVSRBE 2369  
DB 2281 ETEERTKPICVFNWNSILVSGTGSARGCEVVFNRSHVSQCNHMTSPAVLMDVSRBE 2340  
QY 2370 NGBIILPLKLTLYVAAQVTLAALLTFPFLTLRIIRSNQHGIRRNITALGAQLVFLIG 2429  
DB 2341 NGBIILPLKLTLYVAAQVTLAALLTFPFLTLRIIRSNQHGIRRNITALGAQLVFLIG 2400  
QY 2430 INQADLPFACTVYALILHFLYLCFSMALLBAILHYALTEVRDVTNGPMREFTYMLGMGV 2489  
DB 2401 INQADLPFACTVYALILHFLYLCFSMALLBAILHYALTEVRDVTNGPMREFTYMLGMGV 2460  
QY 2490 PAFITGLAVGDPBEGYGNDFCMLSYDTLINSFAGPVAFAVMSVFLYTLAARASCAAQ 2549  
DB 2461 PAFITGLAVGDPBEGYGNDFCMLSYDTLINSFAGPVAFAVMSVFLYTLAARASCAAQ 2520  
QY 2550 ROGFEEKGPVSGLOPSPFAYLLLSATWILALSVNSDPLLPHYLPATCNCIOGPFIFLSY 2609  
DB 2521 ROGFEEKGPVSGLOPSPFAYLLLSATWILALSVNSDPLLPHYLPATCNCIOGPFIFLSY 2580  
QY 2610 VVLSKEVRKALKLACSRKSPDPALTYTSTLTSYNCSPYVADGRLYQPYGDSAGLSHST 2669  
DB 2581 VVLSKEVRKALKLACSRKSPDPALTYTSTLTSYNCSPYVADGRLYQPYGDSAGLSHST 2640  
QY 2670 SRSGKSQPSYIFPLLRBESSALNPGQPGGLGDPGLFLEGDOQHDPDTSDSLSLEBD 2729  
DB 2641 SRSGKSQPSYIFPLLRBESSALNPGQPGGLGDPGLFLEGDOQHDPDTSDSLSLEBD 2700  
QY 2730 QSGSYASTHSSDSRBEEREBEALAPPGQGDMSLLGPABKBLPLHSTYKDGPGGKAP 2789  
DB 2701 QSGSYASTHSSDSRBEEREBEALAPPGQGDMSLLGPABKBLPLHSTYKDGPGGKAP 2760  
QY 2790 WPGDFGTAKSSGNGAPBERLRENGDALSRBSGLPPLGSSAQPHKGLIKKCLPTISE 2849  
DB 2761 WPGDFGTAKSSGNGAPBERLRENGDALSRBSGLPPLGSSAQPHKGLIKKCLPTISE 2820  
QY 2850 KSLILRLPLEOCTGSSRGSASBGRGPPRPPRQSIQOBLQNGMPFAMSTIKAGTUBE 2909  
DB 2821 KSLILRLPLEOCTGSSRGSASBGRGPPRPPRQSIQOBLQNGMPFAMSTIKAGTUBE 2880  
QY 2910 DSSGSEFLFFNFIAH 2923



**RESULT 6**

AAU74826  
AAU74826 standard; protein; 2936 AA.

AAU74826;

23-APR-2002 (first entry)

Human REPT9 protein.

REPT9; human; antiinflammatory; cytoskeletal; immunosuppressive; antiviral; anti-HIV; antiaerobic; anticoagulant; nociceptive; neuroprotective; antiallergic; antibody; immunogen; endometriosis; gastrointestinal disorder; gastritis; oesophageal carcinoma; Crohn's disease; irritable bowel syndrome; ulcerative colitis; endocrine disorder; hypochlorhydria; Kallman's disease; autoimmune disease; inflammatory disease; infertility; receptor; acquired immune deficiency syndrome; AIDS; rheumatoid arthritis; allergy; osteoarthritis; diabetes mellitus; multiple sclerosis; systemic lupus erythematosus; cell proliferative disorder; cancer; developmental disorder; Duchenne muscular dystrophy; Becker muscular dystrophy; neurological disorder; epilepsy; Alzheimer's disease; Huntington's disease; reproductive disorder.

Homo sapiens.

MW000198354-A2.

27-DEC-2001.

21-JUN-2001; 2001MO-US19942.

21-JUN-2000; 2000US-21407P.  
25-AUG-2000; 2000US-22804SP.  
12-DEC-2000; 2000US-255104P.

(INCY-) INCYTE GENOMICS INC.

Giffin JA, Kallick DA, Tritonley CM, Yue H, Nguyen DB, Tang YT, Lai P, Policky JL, Azimzai Y, Lu DM, Grahl R, Yao MG, Burford N, Halalla AJA, Baughn WR, Bandman O, Patterson C, Yang J, Gandhi AR, Warren BA, Ding L, Sanjanwalla MS, Duggan BM, Lu Y;  
MPI: 2002-090432/12.  
N-PDSB; ABK5177.

Ten human receptors (referred to as REPT9-1 to REPT9-12), useful in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell proliferative (e.g. cancer) disorders -

Claim 53; Page 131-138; 157pp; English.

This invention relates to twelve human receptors cDNA sequences referred to as REPT9-1 to REPT9-12), and the proteins encoded thereby. The proteins of the invention may have anti-inflammatory, cytostatic, immunosuppressive, antiviral, anti-HIV, antiaerobic, musclic active general, anticonvulsant, nociceptive, neuroprotective, antiallergic activities. The sequences of the invention may be used to produce REPT9 agonists or antagonists, and the protein sequences may be used to raise anti-REPT9 antibodies. These molecules and the REPT9 polypeptides are polymeric peptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), oesophageal carcinoma, Crohn's disease, irritable bowel syndrome, ulcerative colitis), endocrine (e.g. hypochlorhydria disorder, Kallman's disease), autoimmune/inflammatory (e.g. acquired immune deficiency syndrome (AIDS)), rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus, multiple sclerosis, systemic lupus erythematosus), cell proliferative (e.g. cancer), developmental (e.g. Duchenne and Becker muscular dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,

Query Match	98.3%	Score 15279	DB 23	Length 2936	
Best Local Similarity	98.4%	Pred. No. 0			
Matches 2893	Conservative 2	Mismatches 14	Indels 32	Gaps 8	
Qy 1	MRSPATGVPLPT-PPPLLLLLLLLLPPPLLDGVGFCRSISGRSGSSGACAPMGLCP	59			
Db 1	MRSPATGVPLPTPEPPPLLLLLLLLLPPPLLDGVGFCRSISGRSGSSGACAPMGLCP	60			
Qy 60	SSASNNLWLYTSRCDADAGTELGHVPHHDGLRWCEPSEAHIPLPAPBEGCPSLCRLGI	119			
Db 61	SSASNNLWLYTSRCDADAGTELGHVPHHDGLRWCEPSEAHIPLPAPBEGCPSLCRLGI	120			
Qy 120	GGHLSPGCKTLPEHECLNAPRLRCSCKLAQAPGRAGERSPESISGGRKKNNVTAP	179			
Db 121	GGHLSPGCKTLPEHECLNAPRLRCSCKLAQAPGRAGERSPESISGGRKKNNVTAP	180			
Qy 180	QFPPSPQATVPENQAPGTPVASIRAPIDPBGEAGRLPYTMALFDRSNQFSLDPVTG	239			
Db 181	QFPPSPQATVPENQAPGTPVASIRAPIDPBGEAGRLPYTMALFDRSNQFSLDPVTG	240			
Qy 240	AVTTAEELDEBETSTHYERVTAODHGMRSSALATLLITDPTMDHPVPEQCEKESLR	299			
Db 241	AVTTAEELDEBETSTHYERVTAODHGMRSSALATLLITDPTMDHPVPEQCEKESLR	300			
Qy 300	ENLEVGVEVLTVATDADAPNNAILYRLLEGSGSGSPSEVFEIDPRSGLRTGPPVREE	359			
Db 301	ENLEVGVEVLTVATDADAPNNAILYRLLEGSGSGSPSEVFEIDPRSGLRTGPPVREE	360			
Qy 360	VESYQLTVKASDGRDPGPRSTTAAYVLSVEDNDNAPQSESKRYTVQVREDVTPGAPVL	419			
Db 361	VESYQLTVKASDGRDPGPRSTTAAYVLSVEDNDNAPQSESKRYTVQVREDVTPGAPVL	420			
Qy 420	RVTAASDRDKSNMVHRSIMSGNARGPYLDAQTGALDVPSPLDYETTKETTLKVRADG	479			
Db 421	RVTAASDRDKSNMVHRSIMSGNARGPYLDAQTGALDVPSPLDYETTKETTLKVRADG	480			
Qy 480	GRPLSNVSGLVTVQVLIDINDNAPIFVSTPFOATVLESVPLGYLVHVOAIDADAGNAR	539			
Db 481	GRPLSNVSGLVTVQVLIDINDNAPIFVSTPFOATVLESVPLGYLVHVOAIDADAGNAR	540			
Qy 540	LERLAGVGDPPFTLNGNGWISVAAEIDREEDPFSFVBARHDGTPALITASASVYT	599			
Db 541	LERLAGVGDPPFTLNGNGWISVAAEIDREEDPFSFVBARHDGTPALITASASVYT	600			
Qy 600	VLDVNDNNPTFQPEYTVRLNEDAAVGTSVTVSAVDRDAHSYITVQITSGNTRNRPST	659			
Db 601	VLDVNDNNPTFQPEYTVRLNEDAAVGTSVTVSAVDRDAHSYITVQITSGNTRNRPST	660			
Qy 660	SQSGGGVSLATPLDYKLERQVYLATASGTFODTPAQIYVNTDANTHPPVQSSHYTV	719			
Db 661	SQSGGGVSLATPLDYKLERQVYLATASGTFODTPAQIYVNTDANTHPPVQSSHYTV	720			
Qy 720	NVNEEDRAGCTTVLISATDEDTGENNARIITYFMEDSIPQFRIADDTGAVTTQAEIDYEDQ	779			
Db 721	NVNEEDRAGCTTVLISATDEDTGENNARIITYFMEDSIPQFRIADDTGAVTTQAEIDYEDQ	780			
Qy 780	SYTLATATARDNGIPQKSDTTYTLILVNDVNDNAPQPLRDSYQSGSVYEDVPFTTSVLQISA	839			
Db 781	SYTLATATARDNGIPQKSDTTYTLILVNDVNDNAPQPLRDSYQSGSVYEDVPFTTSVLQISA	840			
Qy 840	TDRSGANGVFTTFOGDDGDDGDFIVESISGIVRTLRRLDRENVAGQVLRAYAVDKMP	899			
Db 841	TDRSGANGVFTTFOGDDGDDGDFIVESISGIVRTLRRLDRENVAGQVLRAYAVDKMP	900			
Qy 900	PAPTEAVTVTVLVDVNDNPPVFEQDEPFDVFEENSPIGLAVAVATATPDGNTAQIMYQ	959			

Db 901 PAATMEVTVTLVDVNDNPEVFEQEDVFEENSPIGLAVARVATADDEGTNAQIMTQ 960  
 QY IVEGNIPEVFOLDIFSGELTALVDLDEDRPEYVLVIQATSAPLVSRAVTHVRLDRND 1019  
 Db 961 IVEGNIPEVFOLDIFSGELTALVDLDEDRPEYVLVIQATSAPLVSRAVTHVRLDRND 1020  
 QY 1020 PVLGNPELIFNNVYTNRSSPPGGAIGRVPAHDPIISLTYSPERGNELSLVLNAST 1079  
 Db 1021 PVLGNPELIFNNVYTNRSSPPGGAIGRVPAHDPIISLTYSPERGNELSLVLNAST 1080  
 QY 1080 GEKLSRALDNNRPLKAIMSVVSDGVHSTYACALAVTITIBMTHTSTITLLEMSPE 1139  
 Db 1081 GEKLSRALDNNRPLKAIMSVVSDGVHSTYACALAVTITIBMTHTSTITLLEMSPE 1140  
 QY 1140 RFLSPILGLFIQAVATLAPDPHVVFNQVQRTDAPGHIILNVSLVSGPPPGGAPP 1199  
 Db 1141 RFLSPILGLFIQAVATLAPDPHVVFNQVQRTDAPGHIILNVSLVSGPPPGGAPP 1200  
 QY 1200 LPSEDLQERLYNRSLLTAISAQVLPFDNICTLREPCENYMCVSVLRFDSAPFIASS 1259  
 Db 1201 LPSEDLQERLYNRSLLTAISAQVLPFDNICTLREPCENYMCVSVLRFDSAPFIASS 1260  
 QY 1260 SVLFRRPIHPVGLRCPCBPFTDGYCETEVDLCYSRPGCHGRCSRREGYTCLCDGYT 1319  
 Db 1261 SVLFRRPIHPVGLRCPCBPFTDGYCETEVDLCYSRPGCHGRCSRREGYTCLCDGYT 1320  
 QY 1320 GEHCESVANSGRCTPGVCKNGCTCVNLLVGFCKDCPSGDFEKPCQVTTSPSPASFTT 1379  
 Db 1321 GEHCESVANSGRCTPGVCKNGCTCVNLLVGFCKDCPSGDFEKPCQVTTSPSPASFTT 1380  
 QY 1380 FRGLRQRFHTLALSPATKERDGLLYNRFNEKHDFVALEVIQEOVLTFPAGESSTTV 1439  
 Db 1381 FRGLRQRFHTLALSPATKERDGLLYNRFNEKHDFVALEVIQEOVLTFPAGESSTTV 1440  
 QY 1440 SPFVPGSVSDGOMHTVOLKYNNPILGOTGLPGSPSOKXAVVTVNDCDTPVALRGSVL 1499  
 Db 1441 SPFVPGSVSDGOMHTVOLKYNNPILGOTGLPGSPSOKXAVVTVNDCDTPVALRGSVL 1500  
 QY 1500 GNYSCAAOGTQGSKKSLDLTGPILLGVPDLPESPVRMRQFVGCMRLQVDSRIHDA 1559  
 Db 1501 GNYSCAAOGTQGSKKSLDLTGPILLGVPDLPESPVRMRQFVGCMRLQVDSRIHDA 1560  
 QY 1560 DFLANNVTGVCRAKKNVCSNCTCHNGGTCVNMDFSCCEPYGFGKSCAQOMANPOHF 1619  
 Db 1561 DFLANNVTGVCRAKKNVCSNCTCHNGGTCVNMDFSCCEPYGFGKSCAQOMANPOHF 1620  
 QY 1620 LGSSLVAMHGLSLPIISOPTWLSLMPRTROADGVTLAITGRSTITLQLRBGHVMLSVG 1679  
 Db 1621 LGSSLVAMHGLSLPIISOPTWLSLMPRTROADGVTLAITGRSTITLQLRBGHVMLSVG 1680  
 QY 1680 TGLQASSLRLPGRANDGDWHHAQALGASGPGHAILSPDYGOQARBNLGPRLHGLH 1739  
 Db 1681 TGLQASSLRLPGRANDGDWHHAQALGASGPGHAILSPDYGOQARBNLGPRLHGLH 1740  
 QY 1740 SNTTGGIIPBPAGVARGFRGCLQGVAVSPTPEGVNSLDPSHBSINVEGCGCLPPCCS 1799  
 Db 1741 SNTTGGIIPBPAGVARGFRGCLQGVAVSPTPEGVNSLDPSHBSINVEGCGCLPPCCS 1800  
 QY 1800 NPCEPANSYCSNDMDSYSCDPPGYGDNCTNVCDLNCEHQSCTKRPSPAHGYTCPCRP 1859  
 Db 1801 NPCEPANSYCSNDMDSYSCDPPGYGDNCTNVCDLNCEHQSCTKRPSPAHGYTCPCRP 1860  
 QY 1860 NYLGAPYCETRIIDPCPRGMWGHPTGFCNCNCDVSKGPDPCNKTSGCHCKENHYRPPGSP 1919  
 Db 1861 NYLGAPYCETRIIDPCPRGMWGHPTGFCNCNCDVSKGPDPCNKTSGCHCKENHYRPPGSP 1920  
 QY 1920 TGLLDCYPTGSLSRVCDPBDGOCPCCKGVIGRCORCNDPFAKYTNNGE-----VNYDS 1975  
 Db 1921 TGLLDCYPTGSLSRVCDPBDGOCPCCKGVIGRCORCNDPFAKYTNNGE-----VNYDS 1978  
 QY 1976 CPRAIEGIMWPRTR-----FGLPAAAPCPKSGF-----GTAVRCDCHRGMLPPLNFNC 2025  
 Db 1979 CPRAIEGIMWPRTR-----FGLPAAAPCPKSGF-----GTAVRCDCHRGMLPPLNFNC 2034

QY 2026 TSITSEIKGFAERLQORNESGLDSGRSOOLALLLNATQHTAGYFGSPVKAYOLATRL 2085  
 Db 2035 TSITSEIKGFAERLQORNESGLDSGRSOOLALLLNATQHTAGYFGSPVKAYOLATRL 2094  
 QY 2086 AHSTORGGLSATQDVHFTENTLKVSGALLDTANKRIMBELIQQTEGGTAMLLQHYEVA 2145  
 Db 2095 AHSTORGGLSATQDVHFTENTLKVSGALLDTANKRIMBELIQQTEGGTAMLLQHYEVA 2154  
 QY 2146 SALAQMHTYISPTITVPNIVISVRLDKGNFAGALPRYELARGQPPDLETTVLLP 2205  
 Db 2155 SALAQMHTYISPTITVPNIVISVRLDKGNFAGALPRYELARGQPPDLETTVLLP 2214  
 QY 2206 ESVFRETPEVVRPAPGAEQOEELARQRRHPELSQGEAAVASVIYRTLAGLPHNYDP 2265  
 Db 2215 ESVFRETPEVVRPAPGAEQOEELARQRRHPELSQGEAAVASVIYRTLAGLPHNYDP 2274  
 QY 2266 DRSILRVPRPIINTPVVISVHDEBELPRALDKPVTVOFRLLETERTYPCVFNHNS 2325  
 Db 2275 DRSILRVPRPIINTPVVISVHDEBELPRALDKPVTVOFRLLETERTYPCVFNHNS 2334  
 QY 2326 ILVSGTGMSARGCEVVRNESHVSCQCNHTSPFVLMDSRR-----NGEIL 2374  
 Db 2335 ILVSGTGMSARGCEVVRNESHVSCQCNHTSPFVLMDSRR-----NGEIL 2394  
 QY 2375 PLKTLTYVALGVTLAALLTFEFLTLRLILRSNQHIGIRNLTALGLAQLVFLGINQAD 2434  
 Db 2395 PLKTLTYVALGVTLAALLTFEFLTLRLILRSNQHIGIRNLTALGLAQLVFLGINQAD 2454  
 QY 2435 LPFACTVAILLHFLYLCFESMALIBALHYALTEVDVNTGEMRFYMLGMVPAFIT 2494  
 Db 2445 LPFACTVAILLHFLYLCFESMALIBALHYALTEVDVNTGEMRFYMLGMVPAFIT 2514  
 QY 2495 GLAVGLDEPGYGNPFCMLSTYDTLVNSPAGVAVASVFLYTLAARASCAORQGE 2554  
 Db 2515 GLAVGLDEPGYGNPFCMLSTYDTLVNSPAGVAVASVFLYTLAARASCAORQGE 2574  
 QY 2555 KKGPSGLQPSFAVLLLSATWLLALSNSDITLLFHYLFATCNCIOGPFIPLSYVLSK 2614  
 Db 2575 KKGPSGLQPSFAVLLLSATWLLALSNSDITLLFHYLFATCNCIOGPFIPLSYVLSK 2634  
 QY 2615 EYRKALKIACSKRPPDALTTSSTYNCPSPYADGRLYQPGDSAGSLHSTSRGK 2674  
 Db 2635 EYRKALKIACSKRPPDALTTSSTYNCPSPYADGRLYQPGDSAGSLHSTSRGK 2694  
 QY 2675 SOPSYIFLILREESALNPGQGPGLGDPGSLFLBGOQDOHDPDTSDBLSLEDOQSGSY 2734  
 Db 2695 SOPSYIFLILREESALNPGQGPGLGDPGSLFLBGOQDOHDPDTSDBLSLEDOQSGSY 2754  
 QY 2735 ASTHSDSEEEEEEBAAPPGEGWDSLLGPABRLPLHSTPDGPGPGKAPWPGDF 2794  
 Db 2755 ASTHSDSEEEEEEBAAPPGEGWDSLLGPABRLPLHSTPDGPGPGKAPWPGDF 2814  
 QY 2795 GTTAKESGNGAPREERLBNNDALSRBSLGPLPSSSAQPHKGLIKKCLPTISKSSLL 2854  
 Db 2815 GTTAKESGNGAPREERLBNNDALSRBSLGPLPSSSAQPHKGLIKKCLPTISKSSLL 2874  
 QY 2855 RLPLEQCTGSSRSGSASGSGRPPRPQSLQBOULNGWPIAMSIIKAGTVDEDSGS 2914  
 Db 2875 RLPLEQCTGSSRSGSASGSGRPPRPQSLQBOULNGWPIAMSIIKAGTVDEDSGS 2934  
 QY 2915 E 2915  
 Db 2935 E 2935  
 RESULT 7  
 ABB11404 standard; peptide; 2560 AA.  
 ABB11404;  
 11-JAN-2002 (first entry)

XX Human FLAMINGO 1 homologue, SEQ ID NO:1774.  
 DE Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoiesis regulation; tissue growth; immunomodulator; actin;  
 KW inhibin; chemokinesis; chemokinesis; thrombolytic; oncogenesis;  
 KW myeloid cell disorder; lymphoid cell disorder; aetasma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiaesthetic; antiaesthetic; haemostatic; antiarteriosclerotic;  
 KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antitumor.  
 XX Homo sapiens.  
 OS WO200157188-A2.  
 FN 09-AUG-2001.  
 PD 05-FEB-2001; 2001WO-US03800.  
 PE 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Dmanac RT;  
 PI WPI, 2001-457740/49.  
 DR N-PSDB; ABA08648.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX Claim 20; Page 179-181; 1963jp; English.  
 PS  
 XX Sequences ABA010981-ABA12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA05574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells

CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 CC  
 CC Sequence 2560 AA;  
 SQ  
 Query Match 86.2%; Score 13399; DB 22; Length 2560;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 2530; Conservative 1; Mismatches 13; Indels 4; Gaps 3;  
 QY 380 STTAAPVLSVDDMDNAPQSEKRYVQVREDVTPGAPVLRVLTASDRDKSNAVHYSIM 439  
 DB 13 STTAAPVLSVDDMDNAPQSEKRYVQVREDVTPGAPVLRVLTASDRDKSNAVHYSIM 72  
 QY 440 SGNARQGYLDAGQALDVSPLDYETTKETTLRVRADGGRPPLSNVSGLVTVQVLDIN 499  
 DB 73 SGNARQGYLDAGQALDVSPLDYETTKETTLRVRADGGRPPLSNVSGLVTVQVLDIN 132  
 QY 500 DNAPIPVSTPQATVLSVPLGYLVHVOALDADAGNARLEVRLAGVGHDPPTINNGT 559  
 DB 133 DNAPIPVSTPQATVLSVPLGYLVHVOALDADAGNARLEVRLAGVGHDPPTINNGT 192  
 QY 560 GMSVAAELDRREVDVPSFGVEARDHGTALITASVSVTVLDVNDNNPTFTQPEYVRL 619  
 DB 193 GMSVAAELDRREVDVPSFGVEARDHGTALITASVSVTVLDVNDNNPTFTQPEYVRL 252  
 QY 620 NEDAAVGTSVTVSAVDRDAHSVTTYOITSGNTRNRSITSQSGGLVSLALPDYCLER 679  
 DB 253 NEDAAVGTSVTVSAVDRDAHSVTTYOITSGNTRNRSITSQSGGLVSLALPDYCLER 312  
 QY 680 QYLAATASDGTRODTQIYVNVTDANTHRPVFQSSHYTVNVNDRPAGTTVVLISATDE 739  
 DB 313 QYLAATASDGTRODTQIYVNVTDANTHRPVFQSSHYTVNVNDRPAGTTVVLISATDE 372  
 QY 740 DTGNARITTFMEDSIPQFRIDATGAVTTOQALDYEDOVSYTLATARDNGIPQKSDTT 799  
 DB 373 DTGNARITTFMEDSIPQFRIDATGAVTTOQALDYEDOVSYTLATARDNGIPQKSDTT 432  
 QY 800 YLETLVNDVNDNAPQFRIDSVQSVYEDVPPPTFSVLOISATDRSGANGRYFTTFCGDD 859  
 DB 433 YLETLVNDVNDNAPQFRIDSVQSVYEDVPPPTFSVLOISATDRSGANGRYFTTFCGDD 492  
 QY 860 GDGDFIVESTSGIYRTLRRLDRRENVAGYLRAYAVDGMPPARTPMEVTVTVLDVNNPP 919  
 DB 493 GDGDFIVESTSGIYRTLRRLDRRENVAGYLRAYAVDGMPPARTPMEVTVTVLDVNNPP 552  
 QY 920 VFEQDEDFVVEENSPIGLAVARVATADPDGNTAQMIOIVBGNIDPEVFOIDIFSGELT 979  
 DB 553 VFEQDEDFVVEENSPIGLAVARVATADPDGNTAQMIOIVBGNIDPEVFOIDIFSGELT 612  
 QY 980 ALVDLDYEDDEBEVYLVYQASAPLVSAATVHNLDRNDNPPVLGNBEILFNNTVTRSS 1039  
 DB 613 ALVDLDYEDDEBEVYLVYQASAPLVSAATVHNLDRNDNPPVLGNBEILFNNTVTRSS 672  
 QY 1040 SFPFGAIGRPADPDSDSLTYSFPERGNEISVLTANASGELKLSALDNNRPLRAIMS 1099  
 DB 673 SFPFGAIGRPADPDSDSLTYSFPERGNEISVLTANASGELKLSALDNNRPLRAIMS 732  
 QY 1100 VLVSDDGVHSTVTAOCALRVITITDEMTLHSTTLRLBDMSPERPLSPILGLTQAVATLAT 1159  
 DB 733 VLVSDDGVHSTVTAOCALRVITITDEMTLHSTTLRLBDMSPERPLSPILGLTQAVATLAT 792  
 QY 1160 PPDHVVVFVNTORPTDAPAGHILNVSLVQCPQPGGPPPLPBBDOERLYNRSILTAI 1219  
 DB 793 PPDHVVVFVNTORPTDAPAGHILNVSLVQCPQPGGPPPLPBBDOERLYNRSILTAI 852  
 QY 1220 SAQGVLPFPDNTICLRBCENYMCVSVTLARFSDSAPFIASSSVLFRPIHPVGGHRCRCPG 1279  
 DB 853 SAQGVLPFPDNTICLRBCENYMCVSVTLARFSDSAPFIASSSVLFRPIHPVGGHRCRCPG 912  
 QY 1280 FTGDYCTEYDLCYSRPCGPHGCRSBRGGYTCLCRDGYTGEHCYVSANSGRCTPGVCNK 1339

```

Db 913 FPGDYCEYEDLCLCYSPGCHGRCSRREGGYTCLCDGYGCEHCEVASASGRCTPGVCKN 972
Qy 1340 GGTGCVLLVGVGFCDCPSGDFEKPVCQVITRSPFASPIITPGRIRORFHTALASPAE 1399
Db 973 GGTGCVLLVGVGFCDCPSGDFEKPVCQVITRSPFASPIITPGRIRORFHTALASPAE 1032
Qy 1400 RDGLLLYNGRPNKEDFVALLVLEIIOEQVOLTFSAGSESTTVSPVPQGVSDGQWHTYOLK 1459
Db 1033 RDGLLLYNGRPNKEDFVALLVLEIIOEQVOLTFSAGSESTTVSPVPQGVSDGQWHTYOLK 1092
Qy 1460 YNKPVLGGTGLPGGBSEOKAVVTVVDCDTGVALRTGSVLGNTSCAAGTGGSKSLDL 1519
Db 1093 YNKPVLGGTGLPGGBSEOKAVVTVVDCDTGVALRTGSVLGNTSCAAGTGGSKSLDL 1152
Qy 1520 TGPLLLGVDPDLPESPVVRKQFVGCGRNLQVDSRHIDMADFANNGTVPGCCAKKNVCD 1579
Db 1153 TGPLLLGVDPDLPESPVVRKQFVGCGRNLQVDSRHIDMADFANNGTVPGCCAKKNVCD 1212
Qy 1580 SNTCHNGGTCVQWMDAFSCECPGLGFGKSCAQEMANPOHFLGSSLYAMHGLSLPISQPMY 1639
Db 1213 SKTCHNGGTCVQWMDAFSCECPGLGFGKSCAQEMANPOHFLGSSLYAMHGLSLPISQPMY 1272
Qy 1640 LSMFTTRQADGYLLQAITRGRSTITLQIRBEGHVM SVBGTGLQASSLRLBGRANDGM 1699
Db 1273 LSMFTTRQADGYLLQAITRGRSTITLQIRBEGHVM SVBGTGLQASSLRLBGRANDGM 1332
Qy 1700 HHQQLMAGSGGFGHAILSFYDQGRABEGLGRLGLHLNNTYVGIKGPAGVARGFR 1759
Db 1333 HHQQLMAGSGGFGHAILSFYDQGRABEGLGRLGLHLNNTYVGIKGPAGVARGFR 1392
Qy 1760 GCLQGVRSVDPBGVNSLDPBSHGESINVBQCSLPDPCDSNPPCANSYCSNDMSYSCSC 1819
Db 1393 GCLQGVRSVDPBGVNSLDPBSHGESINVBQCSLPDPCDSNPPCANSYCSNDMSYSCSC 1452
Qy 1820 DPGYVDCNCTNVCDLNPCEHQSCTRKPSAPHGYTCECPNTYGPICETRIIDQPCRGW 1879
Db 1453 DPGYVDCNCTNVCDLNPCEHQSCTRKPSAPHGYTCECPNTYGPICETRIIDQPCRGW 1512
Qy 1880 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPGSPYCLLDCYPTGSLSRVCDPE 1939
Db 1513 GHPTGCPNCNDVSKGPDPCNKTSGECHCKENHYRPPGSPYCLLDCYPTGSLSRVCDPE 1572
Qy 1940 DGQCPCKPGVIGRQCRCDNPPFAEVTNNGCEVNVYDSCPRAIEMGIMWPRTRFGLPAAAPC 1999
Db 1573 DGQCPCKPGVIGRQCRCDNPPFAEVTNNGCEVNVYDSCPRAIEMGIMWPRTRFGLPAAAPC 1632
Qy 2000 PKGSPGTAVHCDHNGMLPBNLFCNCTSTTRSELKGPABRLQBNESGLDSGRSQOLALL 2059
Db 1633 PKGSPGTAVHCDHNGMLPBNLFCNCTSTTRSELKGPABRLQBNESGLDSGRSQOLALL 1692
Qy 2060 RNATQTAGYFSGDVAVAYOLATRLAHBESTRGFGSATODVHFENLIRVGSALLDPA 2119
Db 1693 RNATQTAGYFSGDVAVAYOLATRLAHBESTRGFGSATODVHFENLIRVGSALLDPA 1752
Qy 2120 NKCHMELIOQTREGTAMLLQHYEAVASALAKQNNRHLYLSPTITVNTNIVISVRLDKNP 2179
Db 1753 NKCHMELIOQTREGTAMLLQHYEAVASALAKQNNRHLYLSPTITVNTNIVISVRLDKNP 1812
Qy 2180 AGAKLPRYALRGEPDLETTVILPESVRETPPVVVRPGRPEABQPEELARQRRHDE 2239
Db 1813 AGAKLPRYALRGEPDLETTVILPESVRETPPVVVRPGRPEABQPEELARQRRHDE 1872
Qy 2240 LSGGEAVASYIIVRTLAGLLPHNYDPDKSLRVPKPIINTPVVSI SVHDEBELPRALD 2299
Db 1873 LSGGEAVASYIIVRTLAGLLPHNYDPDKSLRVPKPIINTPVVSI SVHDEBELPRALD 1932
Qy 2300 KPVTVQPRLLLETERTPICVPMNHSTLUSGTGMSARGCEVFRMESHYSCCCNMTSF 2359
Db 1933 KPVTVQPRLLLETERTPICVPMNHSTLUSGTGMSARGCEVFRMESHYSCCCNMTSF 1992
Qy 2360 AYAMDVSRRENGEILPLKTLTYVALGVTLAALLTPFLTLILNSNOHIGIRNLTAL 2419

```

```

Db 1993 AYAMDVSRRENGEILPLKTLTYVALGVTLAALLTPFLTLILNSNOHIGIRNLTAL 2052
Qy 2420 GLAOLVFLGINADLPACTVIALILHFLYCTPSMALLALHLYRALTEVRDNTGPM 2479
Db 2053 GLAOLVFLGINADLPACTVIALILHFLYCTPSMALLALHLYRALTEVRDNTGPM 2112
Qy 2480 RPYVVLGNGVPAPITGLAVGLDPRGYNDFCWLSTYDITLWSPFAGVAVAVMSVFLYI 2539
Db 2113 RPYVVLGNGVPAPITGLAVGLDPRGYNDFCWLSTYDITLWSPFAGVAVAVMSVFLYI 2172
Qy 2540 LAARSCAORQGRBKGPVSGLOPSPFVALLLSATWLLALLSVNSDTLLPHYLPATNC 2599
Db 2173 LAARSCAORQGRBKGPVSGLOPSPFVALLLSATWLLALLSVNSDTLLPHYLPATNC 2232
Qy 2600 IQGPIFLSYVVLKRVKALKACSRKPSDPALITKSTLTSYNCPSPYADGRLYOPY 2659
Db 2233 IQGPIFLSYVVLKRVKALKACSRKPSDPALITKSTLTSYNCPSPYADGRLYOPY 2292
Qy 2660 GDSAGLSHSTRSGKSPSYIPLLRBSALNPGQPPGLGD -PGSLFLEG--DDQHP 2716
Db 2293 GDSAGLSHSTRSGKSPSYIPLLRBSALNPGQPPGLGD -PGSLFLEG--DDQHP 2352
Qy 2717 DT-DSDSLSTEDDQSGSYASTHSSDSRBRBRBRBRBRBRBRBRBRBRBRBRBRBRBR 2775
Db 2353 XTRDPDSDLSEDDQSGSYASTHSSDSRBRBRBRBRBRBRBRBRBRBRBRBRBRBRBR 2412
Qy 2776 STPKDGPBGPKAPPGDPFGTTAKSSGNGAPBERRLRNGALSRGSLGPLPGSSAOPH 2835
Db 2413 STPKDGPBGPKAPPGDPFGTTAKSSGNGAPBERRLRNGALSRGSLGPLPGSSAOPH 2472
Qy 2836 KGLKKKCLPTISEKSLRLPLEQCTSSSRGSSASBEGSRGPPPPRPSLOBLQNGV 2895
Db 2473 KGLKKKCLPTISEKSLRLPLEQCTSSSRGSSASBEGSRGPPPPRPSLOBLQNGV 2532
Qy 2896 MPAMSTAGTVDEDSGSEFLPRFLH 2923
Db 2533 MPAMSTAGTVDEDSGSEFLPRFLH 2560

RESULT 8
AAB42192
ID AAB42192 standard; Protein, 2405 AA.
XX
AC AAB42192;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1956 polypeptide sequence. SEQ ID NO:3912.
XX
KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
vulnerary; antiapoptotic; antiparkinsonian; neurotrophic; neuroprotective;
anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antineumatic; antithyroid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.

```

PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinketsu RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC76401.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 3067-3072; 5507pb; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteoplastic; anticonvulsant; antiallergic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasoregic;  
 CC antidiabetic; hypotensive; dermatologic; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antihypertensive; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 2405 AA;

Query Match 80.7%; Score 12541.5; DB 21; Length 2405;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 2354; Conservative 23; Mismatches 28; Indels 3; Gaps 2;

QY 516 ESVPLGLVLAHVOAIDADAGNARLEVLAVGVHDPPTINNKGWLSVAELDREYDF 575  
 Db 1 ESVPLGLVLAHVOAIDADAGNARLEVLAVGVHDPPTINNKGWLSVAELDREYDF 60  
 QY 576 YSFGEARADHGTPLATASASVTVLVDNNDNPTFTQPEYTVRLMEDAAVGSVTVASAV 635  
 Db 61 YSFGEARADHGTPLATASASVTVLVDNNDNPTFTQPEYTVRLMEDAAVGSVTVASAV 120  
 QY 636 DRDASVTVTQITSGNTNRPSITSSGGGLVSLALPLDYGLAEQVYLAVALASDGTQDT 695  
 Db 121 DRDASVTVTQITSGNTNRPSITSSGGGLVSLALPLDYGLAEQVYLAVALASDGTQDT 180  
 QY 696 AQIVNVTDANTHRPVFQSSHYTVVNDPRAGITVVLISATDEDTGENARITTFMEDSI 755  
 Db 181 AQIVNVTDANTHRPVFQSSHYTVVNDPRAGITVVLISATDEDTGENARITTFMEDSI 240  
 QY 756 PQFRIDADTGAVTTQAELEIDYEDQVSYTLATARDNGIPQKSDTTYLTLELVNDVNDAPQF 815  
 Db 241 PQFRIDADTGAVTTQAELEIDYEDQVSYTLATARDNGIPQKSDTTYLTLELVNDVNDAPQF 300  
 QY 816 LRDSYQSVYEDVPFTSVLQISATDRDGLNGRVFTYFQGGDDGDFIYESTSGIYRT 875  
 Db 301 LRDSYQSVYEDVPFTSVLQISATDRDGLNGRVFTYFQGGDDGDFIYESTSGIYRT 360  
 QY 876 LRLIDRENVAVQVYLAVALVDKGMPPARTPMEVTVTVLVDNDNPVFEQDEBDFVVEENSP 935  
 Db 361 LRLIDRENVAVQVYLAVALVDKGMPPARTPMEVTVTVLVDNDNPVFEQDEBDFVVEENSP 420

QY 936 IGLAVARVATDPDEGTNAQIMQIVVEGNIPVEVOLDIPSGELTALVDLDYEDRPEYVLV 995  
 Db 421 IGLAVARVATDPDEGTNAQIMQIVVEGNIPVEVOLDIPSGELTALVDLDYEDRPEYVLV 480  
 QY 996 IQATSAPLVSAATVHVALLDRNDNPPVLGNFELLFNNTYTRSSSPFGAAGRPAHPD 1055  
 Db 481 IQATSAPLVSAATVHVALLDRNDNPPVLGNFELLFNNTYTRSSSPFGAAGRPAHPD 540  
 QY 1056 ISDSLTVSPERGNELSVLINAATGELKSRALDNNRPLEAIMSVSDGHSVTAQCAL 1115  
 Db 541 ISDSLTVSPERGNELSVLINAATGELKSRALDNNRPLEAIMSVSDGHSVTAQCAL 600  
 QY 1116 RVITITDMLTHSTITLLEDMSPERFLSPILGLFQAVATLATPDPHVAVNVQRTDA 1175  
 Db 601 RVITITDMLTHSTITLLEDMSPERFLSPILGLFQAVATLATPDPHVAVNVQRTDA 660  
 QY 1176 PGHILNVSLVQGPPEGGPPLBSEDQERLYNRSULTAISARVLPEDNICKRE 1235  
 Db 661 -SSNINVTESALLPGVGRG--QFPSEDLQEOIYNRTLLTITSTORVLPEDNICKRE 717  
 QY 1236 PCENYMCVSVLAPDSAPPLIASSSVLPFRPIHPVGLRCRCPGPTGDYCEVLDLQYR 1295  
 Db 718 PCENYMCVSVLAPDSAPPLIASSSVLPFRPIHPVGLRCRCPGPTGDYCEVLDLQYR 777  
 QY 1296 PCGPHGRCSRREGYTLCRDGYTGEHCYVANSRGCTPGVCKNGGTCVNLVGGFKDC 1355  
 Db 778 PCGANGCRSRREGYTLCRDGYTGEHCYVANSRGCTPGVCKNGGTCVNLVGGFKDC 837  
 QY 1356 PSGBFEKPYQVYTRSPASPTFRGLRPFHTLALSPATXRDLILYNGFENKHD 1415  
 Db 838 PSGBFEKPYQVYTRSPASPTFRGLRPFHTLALSPATXRDLILYNGFENKHD 897  
 QY 1416 FVALEVLQEOVOLTFSSAGESTTVSPVPVGVSDQWHTVOLKYNRPPLGQTLPOGPS 1475  
 Db 898 FVALEVLQEOVOLTFSSAGESTTVSPVPVGVSDQWHTVOLKYNRPPLGQTLPOGPS 957  
 QY 1476 EOKYAVTVTDCDCTGVALARFSGVLGNYSCAAQGTGGSKSLDTGPLLAGVVDLPESF 1535  
 Db 958 EOKYAVTVTDCDCTGVALARFSGVLGNYSCAAQGTGGSKSLDTGPLLAGVVDLPESF 1017  
 QY 1536 PYRRROFVGCMRLQVDSRHIDMADFIANNGTVPGCCAKKNVCDNSNCHNGGTCVNDMDA 1595  
 Db 1018 PYRRROFVGCMRLQVDSRHIDMADFIANNGTVPGCCAKKNVCDNSNCHNGGTCVNDMDA 1077  
 QY 1596 FSCCEPLFGGKSCAQMANNPQHLGSSLVAMHGLSLPISQPMWLSLMPFRTRQADGVLLQ 1655  
 Db 1078 FSCCEPLFGGKSCAQMANNPQHLGSSLVAMHGLSLPISQPMWLSLMPFRTRQADGVLLQ 1137  
 QY 1656 AITRGASTITLQLRBGMVLSVGTGLQASSLKLPERANDGMWHAQLALGASGGFGHA 1715  
 Db 1138 AITRGASTITLQLRBGMVLSVGTGLQASSLKLPERANDGMWHAQLALGASGGFGHA 1197  
 QY 1716 ILSFDYQOQABENGLPRLHGLHSNTYVGIIPGAGVARGFGCLQGVRSPTPEGVN 1775  
 Db 1198 ILSFDYQOQABENGLPRLHGLHSNTYVGIIPGAGVARGFGCLQGVRSPTPEGVN 1257  
 QY 1776 SLDPSSHESINVEQCSLPDPCDNPNCANYSYCSNDMSYSCSDPGYGYDNCNTVCDLN 1835  
 Db 1258 SLDPSSHESINVEQCSLPDPCDNPNCANYSYCSNDMSYSCSDPGYGYDNCNTVCDLN 1317  
 QY 1836 PCRHQSVCTKRPAPAGHYTTECPNTIAGPYCEFRIDOPCPCRGWGHPTCPCCNDVSKGF 1895  
 Db 1318 PCRHQSVCTKRPAPAGHYTTECPNTIAGPYCEFRIDOPCPCRGWGHPTCPCCNDVSKGF 1377  
 QY 1896 DPDCNKTSGSCCHKENHYRPGSPPTCLLCCYPTGSLISRCVDPEDDQCPCKPGVIGQCD 1955  
 Db 1378 DPDCNKTSGSCCHKENHYRPGSPPTCLLCCYPTGSLISRCVDPEDDQCPCKPGVIGQCD 1437  
 QY 1956 RCDNPFALVTTCNCAVNDSCPRALBAGIWPBTRFGLPAAACCPKGSFGTAVAHCDENR 2015  
 Db 1438 RCDNPFALVTTCNCAVNDSCPRALBAGIWPBTRFGLPAAACCPKGSFGTAVAHCDENR 1497  
 QY 2016 GMLPBNLFCNCTITSPBILKGFARLQNBESGLDSGSQLALLLRNATQHTAGTGGSDV 2075

```

Db      1498 GMLPEVLFCTSTSTFSELKGPARRLPDNNSSGLDSGSOQLALLRLAATHTGYPESDVK 1557
Qy      2076 VAVQATRLRLAAHSTORGFASATQDVHFTENILRRGSAALDTPANRHWELIQTEGTA 2135
Db      1558 VAVQATRLRLAAHSTORGFASATQDVHFTENILRRGSAALDTPANRHWELIQTEGTA 1617
Qy      2136 WLLQHYEAVASALAAQMRHTYLSPTTIVTPNIVSVRLDKGFAGAKLPRYALRGEOP 2195
Db      1618 WLLQHYEAVASALAAQMRHTYLSPTTIVTPNIVSVRLDKGFAGAKLPRYALRGEOP 1677
Qy      2196 PDEETVILPESVFERETPPVVRPAGGEAEPEELARRQRHPELSQGEAVASVILYRTL 2255
Db      1678 PDEETVILPESVFERETPPVVRPAGGEAEPEELARRQRHPELSQGEAVASVILYRTL 1737
Qy      2256 AGLLPHNYDPDKASLKVPKAPIINTPVVSVSVADDEELLPRALDKVTVQFRLLEERT 2315
Db      1738 AGLLPHNYDPDKASLKVPKAPIINTPVVSVSVADDEELLPRALDKVTVQFRLLEERT 1797
Qy      2316 KPICVFMNHSILVSGTGMASRCCEVFNESHVSCQCNMTSFAVLMVSRRENGEILP 2375
Db      1798 KPICVFMNHSILVSGTGMASRCCEVFNESHVSCQCNMTSFAVLMVSRRENGEILP 1857
Qy      2376 LKTLTVVAVGVTAAALLTFEFLTLRLIRSNQGIIRNLTALGLAQLVFLGINQADL 2435
Db      1858 LKTLTVVAVGVTAAALLTFEFLTLRLIRSNQGIIRNLTALGLAQLVFLGINQADL 1917
Qy      2436 PFACVIAIILHFLVLCSTSWALLLEALHLYRALTEVDVNTGMRFYMLGMGVPAFITG 2495
Db      1918 PFACVIAIILHFLVLCSTSWALLLEALHLYRALTEVDVNTGMRFYMLGMGVPAFITG 1977
Qy      2496 LAVGLDPEGYGNDFCWLSTYDTLINSFAGVPAFAVSMGVFLYTLAARASCAAQRGFEX 2555
Db      1978 LAVGLDPEGYGNDFCWLSTYDTLINSFAGVPAFAVSMGVFLYTLAARASCAAQRGFEX 2037
Qy      2556 KGPVSGLOPSFAVLLLSATMTALSVNSDTLPHYLPATCNCIGCPPLISVYVLSRE 2615
Db      2038 KGPVSGLOPSFAVLLLSATMTALSVNSDTLPHYLPATCNCIGCPPLISVYVLSRE 2097
Qy      2616 VRKALKLACSRKSPDPALTTKSTLTSSVNCPSPYADGRLYQPYGDSAGLSHSTRSGKS 2675
Db      2098 VRKALKLACSRKSPDPALTTKSTLTSSVNCPSPYADGRLYQPYGDSAGLSHSTRSGKS 2157
Qy      2676 QPSYIFPLLRBSBALNPGQPGPLGPGSLFLEGQOQHDPDSDSLBDDOGSATA 2735
Db      2158 QPSYIFPLLRBSBALNPGQPGPLGPGSLFLEGQOQHDPDSDSLBDDOGSATA 2217
Qy      2736 STHSSESESESESESESESESESESESESESESESESESESESESESESESESESESE 2795
Db      2218 STHSSESESESESESESESESESESESESESESESESESESESESESESESESESESE 2277
Qy      2796 TTAKESSESESESESESESESESESESESESESESESESESESESESESESESESESE 2855
Db      2278 TTAKESSESESESESESESESESESESESESESESESESESESESESESESESESESESE 2337
Qy      2856 LPIEQTGSSRGSSABSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 2915
Db      2338 LPIEQTGSSRGSSABSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 2397
Qy      2916 FLFFNFLH 2923
Db      2398 FLFFNFLH 2405

```

```

RESULT 9
AAU68533
ID AAU68533 standard; Protein; 3014 AA.
XX AAU68533;
XX
XX 16-JAN-2002 (first entry)
XX
XX Human novel cytokine encoded by cDNA 790C1P2C_4 #1.
DE

```

```

XX      Human; cytokine; cell proliferation; cell differentiation;
KW      antiinflammatory; stem cell growth factor; activating; inhibin; cancer;
KW      nervous system disease; Huntington's disease; Alzheimer's disease;
KW      Parkinson's disease; Huntington's disease; spinal cord disorder;
KW      head trauma; stroke; myeloid cell disorder; lymphoid cell disorder;
KW      platelet disorder; thrombocytopenia; stem cell disorder;
KW      aplastic anaemia; tissue regeneration; wound healing; ulcer;
KW      osteoporosis; osteoarthritis; bone degenerative disorder;
KW      periodontal disease; fibrosis; reperfusion; immune disorder; SCID;
KW      severe combined immunodeficiency; infection; autoimmune disorder;
KW      multiple sclerosis; rheumatoid arthritis; diabetes mellitus; allergy;
KW      asthma; coagulation disorder; haemophilia; sepsis; nephritis;
KW      inflammatory bowel disease; food supplement; immunogen.
XX
OS      Homo sapiens.
XX
XX      WO200175093-A1.
XX
XX      11-OCT-2001.
XX
XX      30-MAR-2001; 2001WO-US10484.
XX
XX      31-MAR-2000; 2000US--0540217.
XX      23-AUG-2000; 2000US-0649167.
XX      22-SEP-2000; 2000US-0668680.
XX      23-OCT-2000; 2000US-0695618.
XX      30-NOV-2000; 2000US-0728711.
XX      14-MAR-2001; 2000US-0728711.
XX
PA      (HISE-) HISEQ INC.
XX
PI      Tang YT, Asundi V, Zhou P, Xue AJ, Ren P, Zhang J, Wang J, Xu C;
PI      Yang Y, Zaho QA, Chen R, Wang D, Goodrich RW, Liu C, Drmanac RT;
XX
XX      MPI: 2001-626432/72.
XX      N-PDB; AAS59825.
XX
DR      New polypeptides and nucleic acids, useful for diagnosis, treatment of
PT      inflammatory, autoimmune, neurological, myeloid or lymphoid cell, bone
PT      degenerative disorders, cancer and promoting wound healing
XX
PS      Claim 20; Page 249-254; 336pp; English.
XX
XX      The invention relates to isolated human polypeptides (which may be
XX      cytokines) and the polynucleotides encoding them. The proteins are useful
XX      for identifying a compound which binds to it (e.g. modulators, agonists
XX      and antagonists). The polynucleotides are useful as an array for mismatch
XX      detection. The proteins and nucleic acids are useful as nutritional
XX      sources or supplements. The protein exhibits activity relating
XX      to cytokine, cell proliferation, cell differentiation, antiinflammatory,
XX      stem cell growth factor activity, immune stimulating or immune
XX      suppressing and activin or inhibin related activities. The proteins (and
XX      antibodies raised against them) and nucleic acids are therefore useful in
XX      the diagnosis and treatment of diseases and disorders such as cancer,
XX      central and peripheral nervous system diseases and neuropathies,
XX      Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX      lateral sclerosis, spinal cord disorders, head trauma, cerebrovascular
XX      diseases, stroke, myeloid or lymphoid cell disorders, platelet disorders,
XX      thrombocytopenia, stem cell disorders, aplastic anaemia, for
XX      regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
XX      growth, and in tissue repair, healing of burns, incisions, ulcers, for
XX      treating osteoporosis, osteoarthritis, bone degenerative disorders, or
XX      periodontal disease, lung or liver fibrosis, reperfusion injury in
XX      various tissues, various immune deficiencies and disorders including
XX      severe combined immunodeficiency (SCID), bacterial or fungal infections,
XX      autoimmune disorders (e.g. multiple sclerosis, rheumatoid arthritis,
XX      diabetes mellitus, myasthenia gravis), allergic reactions and conditions,
XX      such as asthma or other respiratory problems, coagulation disorders,
XX      haemophilia, septic shock, sepsis, arthritis, nephritis and inflammatory
XX      bowel disease, viral infection and are useful in altering bodily
XX      characteristics. The present sequence represents a novel protein of the
XX      invention.

```



XX Sequence 3014 AA;  
 Query Match 56.3%; Score 8754; DB 22; Length 3014;  
 Best Local Similarity 55.8%; Pred. No. 0;  
 Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;

QY 12 TTPPLLLLLLLLLLPPPLGGDYGPCRSISGRKSSGACAPMGICPSSASNLMLYTSR 71  
 DB 57 TTRAPRELL-----DVGDRGRLAGR-RRVSGAGRP---PLQVRLV----- 93  
 QY 72 CRDAGTELGHVPHHDLRWVCPSEBAHPLRPABEGCPMSCRLIGGHLSPGKLT 131  
 DB 94 ABAFAPLASRL-----RARKLP-----GCAPARLCGTARLC--GALCF 133  
 QY 132 P-----BEAPCLKA-----PLRCOS-----CKLAQAP- 155  
 DB 134 PVGGCAAAQSHALAAPTLLPACRCPPRRPPRCGPICLPFGSGRLRLCALARAAGA 193  
 QY 156 LRAQ-----EASPEESLGGRRKRVNTPA---QFQPSYQAT 189  
 DB 194 VAVGALBAATAGTSPSASPSPPLPPLPPEARAGPARARRGTSGRGLKEPMENYQA 253  
 QY 190 VENQDPAGTPVASLRAIDPDEBAGRLXYMDALPDSRNOFPSLDPVTGATTAABELDR 249  
 DB 254 LPENEAGTLILQHLAHYTTBGBBEKVSYTMGLPDEKRGYFRIDSATGAVSTDSVDR 313  
 QY 250 ETKSTHVFVTAQDHGMPPRSALATLTLLVTDNDHPVFEQOEKESLREMLFVGYEVL 309  
 DB 314 EKEKTHVLKVKAVDSTPRSATYTYTVLVOTNDHSPVFEQSEKREKREMLFVGYEVL 373  
 QY 310 TYRATDGDAPVPAVNTLLYRLLESGSGSPSEVFEIDPDSGVIKTRGPVDRBEVESYQLTYEA 369  
 DB 374 TIRADDRSPRIANLRKYLGGAM---DVFLNBSGGVSTRAVLDRREAAEQLLVEA 429  
 QY 370 SPOGDRPGRSTTAAYVFLSVEDNDNNAPOFSKRYYVQVREDEVYGAPLYRTAASDRKX 429  
 DB 430 NIOGRPGSLATATVYLEVEDENDYPOFSQONTVQVDEVGLNTAALRAQVADRQX 489  
 QY 430 SNAVVHYSIMSGNARQFYLDAQTGALDVVSPLDYETTKETYLRAVRAODGGRPLSNVSG 489  
 DB 490 QNAALHYSILSGNVAQFYLHSLGSLDYNLPDFEDVQKSLSLKAQCGGRPLINSVG 549  
 QY 490 LVTVOVLINDNAPLFVSTPFOATYLSVPLGLVLFVQALDADGNALEFRLA---- 545  
 DB 550 VVSVOVLVDNDEPIFVSSPFOATYLENVPGLGPVPHIQAVDADGSENAKRLHYRLVDEAS 609  
 QY 546 ---GVGH-----DEPTINNGTGMISVAABLDREEDVFGAGVAPRHGTALTA 592  
 DB 610 TFLGGSGAPKAPAPLPDPFPOIHNSSGITVCAELDREVEHYSIGVAADVHGSPMSS 669  
 QY 593 SASVSVTVLVDNNDNFTQPEYTVRLNEDAAVGVSVVAVSVDADHAGSVITYQITSGNT 652  
 DB 670 STSVSITVLDVNDNVPVFTQPTTELRINDAAVGSVTLQARBDANSVITYQLTGNT 729  
 QY 653 RNRFSITSGSGGLSLALPLDYKERQYVLAATVMSDGTRODTAQIVNNTDANTHRPVF 712  
 DB 730 RNRFLSSQGGGLTLALPLDYKOEQYVLAATVMSDGTSHAHVLIIVNTANTHRPVF 789  
 QY 713 GSSHTVANNEDRPAGTIVLLSATDEDEGENARITYEMDSIPORIDADGAVTQAE 772  
 DB 790 QSSHIVSVSEDRPVGTSTATTSTANDEDEGENARITYIADPVQGRIDPDGNTYME 849  
 QY 773 LDYEDQVSTTLAITARNDGIPKSDITTYLEILVINDVNDNAPOFLDSYQSYEDRPPT 832  
 DB 850 LDYENQVAVTLTIMAGDNGIPKSDITTYLEILILANDNNAPOFLMDPVYSGIFEDAPST 909  
 QY 833 SYLQISATRDUSGLANGRVYTTTQGGDDGCDTIVESTSGIVTTLRLDRENAQVLYRAY 892  
 DB 910 SILQYSATRDUSGPNRLLYTFQGGDDGDDPYIFBTSQVIRTORRLDRENAVNLNAL 969  
 QY 893 AYDKGMP--PATTMEVTVVLDVNDNPVVFODEDFVFEENSPIGLAVARTATDPDEG 951

DB 970 AVDRSGPTPLSASVEIQVTIINDNAPMEKDEBELFVEERNNPVGSVAKIRANDPDEG 1029  
 QY 952 TNAQIMQIVEGNIPVEFQLDIPSGELTALVDLYEDRPEVYLVLOATSAPLVSRATVAV 1011  
 DB 1030 PNAQIMQIVGDMKHFHQDLNDLGRAMVELDEPVREYLVVQAASAPLVSAATYHI 1089  
 QY 1012 RLDRNDNPVLVGNFELLFNNYVTRNSSPPGAGIARVPANDPDISLSTYSFERGNEIS 1071  
 DB 1090 LLVQNDNRPVLPVFOQLFNNYVTRNKSNSPFTGVYGCIPADHPVDSISANTYFVQGNELR 1149  
 QY 1072 LVTLNASTGELKLSALDNNRPLRAIMSIVSDGVSTTAQALRYTITDEMTHSTL 1131  
 DB 1150 LLLLDPAIGELQSLKDLNDRPLBALMEVSVDGHSHTATCLKRVITITDMLNSTITV 1209  
 QY 1132 RLEDMSPERFLSPILGLFICAVAATLATPDPHVAVFVNOQBDTDAFGHILNVSVGQPP 1191  
 DB 1210 RLENNSGEKFLSPILALFVEBVAVALSTTKDDVVFVNOQNTDV--SSILNVTSTALLPG 1268  
 QY 1192 GPGGPPPLPSEDDIQLRLYLRSLITLISAQVLPEDDNLCLREPCENYMRCSVLYRPS 1251  
 DB 1269 GVRG--QFPSEDDIQLBOIYLRRLTTLTSTQRVLPFDNLCLEBCEYMRCSVLYRPS 1326  
 QY 1252 SAPFLASSVLPRLPHVGRRCRCRPGFTGDCETEDVLCYSRPGSHGRCSREGGYT 1311  
 DB 1337 SAPFLSSTVLPRLPHVGRRCRCRPGFTGDCETEDVLCYSRPGSHGRCSREGGYT 1386  
 QY 1312 CLCRDGYTGEHCEVASASGRCTPGVCKNGGTCAVLLVGFCKDCDPGDFEKPQCVTRRS 1371  
 DB 1387 CECFEDPFGHECEVDASGRCAVCGKNGGTCAVLLVGFCKDCVPREYEPYCEVTRRS 1446  
 QY 1372 FPAHSFTFRGLRQRFHTLALSPATYERDGLILYNGRFNEKHPVLALEIYQVOVLFS 1431  
 DB 1447 FPPQSFVYFRGLRQRFHTLALSPATYERDGLILYNGRFNEKHPVLALEIYQVOVLFS 1506  
 QY 1432 AGESTTTSVPVPGVSGQWHTVQKYNKPELGGQGLPGCBQKAVAVTVVQGCDDTV 1491  
 DB 1507 AGESTTVPVPGVSGQWHTVQKYNKPELGGQGLPGCBQKAVAVTVVQGCDDTV 1566  
 QY 1492 ALRFGSVLGNTSCAAGTGGSKSLDTLGPPLLGAVDLPESPPVNRQFVGCARNLQV 1551  
 DB 1567 AVRGKIDIGNSCAAQGTQGSKSLDTLGPPLLGAVDLPEDPPVNRQFVGCARNLQV 1626  
 QY 1552 DSRHIDNADPIANNGTVPGPCAKKQVCDSTCNNGGTCAVNOVDAFSGCEPLAFGGKCAQ 1611  
 DB 1627 DGKVDNAGFIANNGTGEGCAARRNFCGRRCQNGGTCAVNMWNYLCECPRLFQKCEQ 1686  
 QY 1612 BMANPOHLSGLVAMHGLSLPISOPWTLSPMPTROADGVLALITRGRSTITLQLEB 1671  
 DB 1687 AMPHQLPFGSGSVWSMDLITLISVPWTLGLMFRIRKSDVLMERTSGGPTSRFQILNN 1746  
 QY 1672 HVMLSVEGTGLQASLSRLRPGRANDGMHHAQLAL---GASGGPGHAI--LSFDYGOQRAE 1727  
 DB 1747 YLQFVSHSGPDBSVNMLSGRLVTDGEWHLLILKXVKEDSEMKHLVTMTLQGMQNK 1806  
 QY 1728 GNLPRLHGLHLSNITVCGILPGAGVYAKRGFGLQGYRVSDTPEGVNSLDPHSGESLNV 1787  
 DB 1807 ADIGKMPGLTVRSVVVVGASLEDKYSVYRGRFGCMQGRMGGTPTNVAATLMMNNALKRV 1866  
 QY 1768 BQGSILPDPQSNCPANSYCSNMDYSGSCDGYGYDNCCTNYCDLNPCHGHSCTPKP 1847  
 DB 1867 KDQCDVDDPCTSSCPDPSRCHDAMEDYSVCVCDKGYLAGINCVDACHLNPCEMGAQVSP 1926  
 QY 1848 SAPHGTYCEPNTYAGPYCETRIDOPCRGMWGHPTGCPNCNDVSKGFPDCKNTSGECH 1907  
 DB 1927 GSPQGYVCEGSHYGPYCEKDKLPCRGWMGNVPGPCHCAVSKGFPDCKNTNGCQ 1966  
 QY 1908 CKENHYRPGSPITGLCDCTYTGSLSRVCPBEDQCCPKRVYTRQDCRCNPPAFVTTN 1967  
 DB 1987 CKENYKLLAODTLPCDDCFPHGSHSRTCDMATQCAKCPGVIRQCRCNPPAFVTTL 2046  
 QY 1968 GCEVYVDSCPRAIRAGIIMPTRFGLPAAAPCPGSGFGTARHCDERGMPLPMLFNCTS 2027  
 DB 2047 GCEVYVDSCPRAIRAGIIMPTRFGLPAAAPCPGSGFGTARHCDERGMPLPMLFNCTS 2106



```

QY 2028 ITFSEIKGAEERLOENESGLDSRGOOLALLLNATQHTAGYFGSDVKVAYOLATLNLH 2087
DB 2107 ISPVDRANNEKLSRETOVDGARALQVRLARSAOHQHTLFGNVRARAYQLGVLQH 2166
QY 2088 ESTQRFGLSATVDHFTENLELVSGALLDTANKRMWELIQTEGGTAMLLQHTYAYASA 2147
DB 2167 ESQOQFDLAATODAFHEDEVINHSGLALLAPATRAWEQIQSEGGTAOLRLRLBEGYFN 2226
QY 2148 LQNMHTLSPPTIYTPNIVISVRLDGNFAGALPRVYLRGQRPDLFTVLPSS 2207
DB 2227 VANNVRRITLRPVITANNITLAVDIFDKFNFTGAVPFRDTHIEFPRLSSVSFPD 2286
QY 2208 VFR---ETPVVVRPA-----PGEAQBPEELARORRHPELSQGEAVAVIYRT 2254
DB 2287 PFRPPEKGPILRAPGRRTPTQTRPGQTEREARISRRRRPDQAQPAVALVIYRT 2346
QY 2255 LAGLPHANTDPDKRSIRVPRPIINPVVISVHDEBELPRALDKPVTVQPLLEBER 2314
DB 2347 LQQLLEPERYDPPRRSIRLPRPIINPMVSTLVYSGAELPRLEPRVLEFALLEBER 2406
QY 2315 TKPICVFMHSLIVSGTGSARGCEVFRNESHVSCOCNHTMSFVAVLMDVSRBNGEIL 2374
DB 2407 TKPVCFVFMHSLAVSGTGSARGCELSRNKTHVACQSHNASFAVLMDSIRREGEVL 2466
QY 2375 PLKTLTYVALGVTALALLTFPFLTLRLRSNQHGRIRNLTALGLAQVLLGLINQAD 2434
DB 2467 PLKITYVAASVSLALVAVLISLVRLSRMLSHIHGMLAVLPSQLVIVIGINQRE 2526
QY 2435 LPPACTIVAILHLFVLTCTPSNALLBALHLYRALTVRDVNTGPMKPYTMLGKVPDFT 2494
DB 2527 NPFLCTVVAIILHYITMSTFAMTLVSVLTVYVRLTEVRNIDTGPMPFYVVGIGIPALIT 2586
QY 2495 GLAVGLDPGSGYGNPCWLTSTVYTLTMSFAGPAPAVMSVFLYILAARSCAQQQRE 2554
DB 2587 GLAVGLDPGSGYGNPCWLTSTVYTLTMSFAGPAPAVIITNTVSVLASVSCQRKHRYG 2646
QY 2555 KKGVPVSGLOPSPAVLLLSATWMLALISVNSDTLLFHYLPATCNCIQGFPIFLSYVLSK 2614
DB 2647 KKGIVSLKTAFLILLISATWMLGLAVNRDLSPHYLPAPSGIGPFLVILPHCVLQ 2706
QY 2615 EYRKALK-LACSKKEPPDALTKST-LTSVNCBPYADG--RLYQPYGDAAGLSHRS 2670
DB 2707 EYRKALKGVLAGKRLLEDSATTRATLLTRSLNCNTTFDGGPDMLTDLGSESTASLDSIV 2766
QY 2671 RS-----GKSQPSYIPFLAREBSALNPGQPPGKDPGSLFLBQDDQ 2713
DB 2767 RDBGIQKLGVSGLVNGSHGRPASLMPNSCD-----PRG----- 2802
QY 2714 HDPDTDSDDLSTEDQSGSYASTHSSDEEBEERBEAAPPQEGWMDSLGPGAERLP 2773
DB 2803 --HDSDSDBELSL-DRQSSSYASHSSDSRDDGVGAERK-----MPARG-----A 2845
QY 2774 LHSTPR-DGPGRGKAPWPDDECTYAKESGNAP-----EERLRNGDAL 2818
DB 2846 VHSPTPGADVANNVPRGWPQDSLAESDSBDPSKPLKVETKVSVELAREEQSHRGKRP 2905
QY 2819 SREGSGAPLPGSSAQP---HKGILKKKCL---PTISEKS---SLRLPLEQCTGSSRSGS 2869
DB 2906 PDESSEGAARLASQPPRQKGLKKKVVYPPVLTITBQTKGRLABKALDEQSFSTSR 2965
QY 2870 ASBGRSGRP---PPRPPRQSLQEQNLGVMPJAMSIKAGTVDEDSGSE 2915
DB 2966 TSSLSGSGPDCATVTKSPGREPRDHLNGV---ANNVTRGSAQADGSDSE 3012

```

```

XX Seven-pass transmembrane receptor-like protein, MEM1.
DB Seven-pass transmembrane receptor, MEM1; therapeutic; diagnostic; MEM2;
XX human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;
KW female reproductive health; lung disorder; brain disorder; schizophrenia;
KW heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;
KW cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;
KW vision-related disorder; neoplastic pathology; MEM7; MEM8.
XX Homo sapiens.
OS WO20014473-A2.
XX 21-JUN-2001.
XX 14-DEC-2000; 2000WO-US33909.
XX 14-DEC-1999; 99US-0170564.
XX 27-DEC-1999; 99US-0173165.
XX 27-DEC-1999; 99US-0173362.
XX 29-DEC-1999; 99US-0173544.
XX 04-JAN-2000; 2000US-9965564.
XX 09-AUG-2000; 2000US-0223929.
XX 13-DEC-2000; 2000US-9965565.
XX (CURA-) CURAGEN CORP.
PI Spaderma SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;
XX WPI; 2001-398154/42.
XX N-PEDB; AAS06332.
XX Novel polypeptide comprising members of protein families (e.g.,
PT seven-pass transmembrane receptor proteins) according to presence of
PT domain and sequence relatedness are useful for treating or preventing,
XX e.g., Alzheimer's and Parkinson's -
XX Claim 1; Fig 2; 162pp; English.
XX The sequence represents the amino acid sequence of seven-pass
XX transmembrane receptor-like protein, MEM1, selected from a group (MEM1-
XX MEM8) comprising members of protein families according to the presence of
XX domain and sequence relatedness, e.g., seven-pass transmembrane receptor
XX protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein
XX (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-
XX MEM8). The MEM polypeptides (I), nucleic acids (II), and antibodies (III)
XX are all useful for treating or preventing a pathology associated with (I)
XX comprising administering (I), (II), or (III) to a subject (preferably a
XX human). In addition, (I), (II), and (III) may be used to manufacture a
XX medicament for treating a syndrome associated with a human disease that
XX is associated with (I). Furthermore, (I) may be used to identify agents
XX that bind to it, screen modulators of its activity and determine the
XX presence or predisposition to a disease associated with altered levels of
XX (I). Disorders for MEM1 include Alzheimer's or Parkinson's disease,
XX cancer, nephrology, and female reproductive health. Disorders for MEM4
XX include those involving the lung and/or brain (e.g., schizophrenia). For
XX MEM5, disorders include heart (arrhythmic disorders) and other muscular
XX disorders, clotting deficiencies and cobalamin deficiencies (e.g.,
XX pernicious anaemia). Such disorders for MEM6 include diabetes, whereas
XX disorders for MEM7 and MEM8 include vision-related disorders, cancer,
XX and other neoplastic pathologies.
SQ Sequence 3014 AA;

```

```

Query Match 56.3%; Score 8754; DB 22; Length 3014;
Best Local Similarity 55.8%; Pred. No. 0; Mismatches 687; Indels 240; Gaps 39;
Matches 1703; Conservative 420;
QY 12 TPEPPLTLLTLLTLLPPLLDQVGPCKRSIGSRGSSGACAPMGWLCPSASNTMLYTSR 71
DB 57 TPRAPREL-----DVGRRDGLAGR-RVSGAGRP-----PLQVRLV----- 93

```

QY 72 CRDAGTELGHVPHHDGLRVWCPBSEBAHPLPPEAGCPCWCRLLGIGCHLSPOGKTLTL 131  
 DB 94 ARSAPALSRRL-----RARTHLP-----GCGARALGCTGARLC--GALCP 133  
 QY 132 P-----BEHPCLKA-----PRLRCOS-----CKLAAPG- 155  
 DB 134 PVEGCGAAGHSAALAPPTILPACRCPPRRPRCPGRCICLPBGGSVRLRLCLALRRPAAQ 193  
 QY 156 LRAG-----ERSPEESLGRKRKNVNTAP-----QFOPPEYQAT 189  
 DB 194 VRVGLALBAATAGTPASPPSPPLPPNLPPEARAGPARARAGTSGRSLKPFMPMYQVA 253  
 QY 190 VPHNOCAGPVPASLRAIDPEBAGALEYTMALPFSKNQFSLDPPTGAVTTAEELDR 249  
 DB 254 LPENEBAGTLILQLAHNYTIEGEEBRSYMEGLPERSRGYFRIDASGAVSTDSVLDR 313  
 QY 250 ETKSTVFVRYTADHGMPPRSALATLITLVTDNTNDHPVEEOOEYKESLREMLEVEYEL 309  
 DB 314 ETKETHTLVAKVADVSTPRSAITTYITLVKQDINDHSVPVEBSETEREVRREMLEVGYEVL 373  
 QY 310 TYRATDGDAPPMANILYRLBSGSGSPSEVFEIDPPSGVIRTRGPVDRREBESYQLTVEA 369  
 DB 374 TIRASDRDSPINAMLRVYGAM-----DVFQLEBSGVSTRVAVLDREBAARYQLLVA 429  
 QY 370 SNOGRDPRGRSTTAAYFLSVEDDNDNAPOPSEKRYVQVEDYTPGAPVLTATYASDRKG 429  
 DB 430 NOGRNPGPLSATYATYIEVEDENDNYPQPSBONVYVQVEDYGLNTAVLTAVQATDRDQ 489  
 QY 430 SNAVYHSIMSGNARGOFLADQTAGLDVYSPLDVETTEKBYTLRYAORCGRPPLSNVSG 489  
 DB 490 QNBAIHYSILSGNVAQFYLHSLSGLLDYINPLDEVDOKYSISTIAQOGGPPILNSBG 549  
 QY 490 LVTVOYLINDNAPIFVSTPPOATVLESVPLGYLVAHQALDADAGNARLEYRLA--- 545  
 DB 550 VVSVOQLDVNDNEPIFVSSPPOATVLENVPLGVPVHIOAVDADSEGNARLHYRLVDIAS 609  
 QY 546 ---GVGH-----DPPFTINNGTGMISVAABELDREEDVPTFGVZARHGHGTPALTA 592  
 DB 610 TFLGSGSAGPKNPAPTPDPFQIHNSSGMITVCAELDREBVERHSYGVAZVHGHGPPMS 669  
 QY 593 SASVSVTVLVANDNNPTFQPEYTVALNEDAAVGTSVVSAVDDASHVITYOITSGNT 652  
 DB 670 STSVSITVLVDVNDNPVFTQPTTELRLNDAVAGSSVLTLOAKRDANDANSVITYOITGANT 729  
 QY 653 RNRFSITSGGGGLVSLALPLDYKLERQYVLAVTASDGTRODTAOIVANVTANTANTHPVF 712  
 DB 730 RNRPALSGRGGGLITLALPLDYKOEQYVLAVTASDGTSHHTAVLINVNTANTANTHPVF 789  
 QY 713 QSSHVTWANNEDRPACTIVVLSATDEDEGENARITTYMEDSIPOPRIDADTGAVTQAE 772  
 DB 790 QSSHVTVSSEDRPVGTSTATLSANDEDEGENARITTYVIODPVPOFRIDDPDSGTMYTMC 849  
 QY 773 LDYEDQVSTYLAITARDNGIPOKSDPTTYLILVNDVNDNAPOFLDSYOGSVYEDVPPT 832  
 DB 850 LDYENQVAYTTLTLMODNGIPOKSDTTTLEILILANDNAPOLMDPTQGSIFEDAPST 909  
 QY 833 SVLQISATDRDSGLNGRVYTFQGGDDGDDPIVESTSGIVTTLRLRLRENAQYVLAAY 892  
 DB 910 SILQVATDRDSGPNRGLLYTFQGGDDGDDFYIPTSGVIRTRORLREBENAVVNLNAL 969  
 QY 893 AYDKMP- PARTPMVYTVVLDVNDNPPVFEDEDEDVVEENSPGLAVARTATDPBG 951  
 DB 970 AYDRSPPLPLASVBEIQTITLINDNAPFEKDELELTVRENNPVGSVVAKIRANDPBG 1029  
 QY 952 TNAQIMYOIVEGNIEVEFQJDI FSGELFALVLDYEDRPEYVLVQATSAPLVSATYAV 1011  
 DB 1030 PRAQIMYOIVEGDMHFRQJDLNGLDAMWELDEVRREYVLVQATSAPLVSATYAVHI 1089  
 QY 1012 RLDRNDNPPVLGNFELLFNNTVYTRSSSPFGAIGRVAPADPDISLITYSFFERGNLS 1071  
 DB 1090 LLDVNDNPPVPLPDEFQILFNNTVYTRKSNFPPTGVIGCI PAHDPVDSILNTYFVGNELR 1149  
 QY 1072 LVLNASTGELSLGSLALDNNRPLEAIMSVLSDGVHSTYAOQALAVTITITDEMLTHSITL 1131

DB 1150 LILLDPAQELQLSHDLNDRPLEALMEVSDGHSYAFCTLRVYITITDMLTNSITV 1209  
 QY 1132 RLEDMSPRRFSLPLGLPQVAAATLAPDPHVVPVNOQDPDAGCHILNLSVGGP 1191  
 DB 1210 RLNNNSOEKFLSPLALFVVGAAVLSSTKDDVFVFNQNDTV--SSNLTNTFALPLFG 1268  
 QY 1192 GPGGPPPLSEHDEORLYNRLSLTASAOVLPEPDNI CLAREPCENYMRCSVYLRFDS 1251  
 DB 1269 GVRG--QFPSESDLOEQYILNRITLITITSTGRVLPFDNICLAREPCENYMKCVSLRFPDS 1326  
 QY 1252 SAPPYASSSVLFRPIHPVGLRCRCPPGFTGDCYCEVLDLCYSPDCGHRGRSREGYT 1311  
 DB 1327 SAPPYASSSVLFRPIHPVGLRCRCPPGFTGDCYCEVLDLCYSPDCGHRGRSREGYT 1386  
 QY 1312 CLCRDGYTGHCEVARSRGCRPGVCKRGKCVNLLVGGFPCCDGSPPREKYCOVTRS 1371  
 DB 1387 CECEDEFTGEHEVDANKRGCANVGCKRGKCVNLLIGFHCVCPPGSEYERPECEVTRS 1446  
 QY 1372 PPAHSFITERGLRORFHTLALSPATKERDGLLYNGRFNEKHDFVAALEVIQEOVLTFPS 1431  
 DB 1447 FPQGSFVTFRGLRQRFHTTISLTPATQBERNGLLYNGRFNEKHDFVAALEVIQEOVLTFPS 1506  
 QY 1432 AGESTTYSPTVPGGVSDQMHVYQLKTYNPLSGTGLPOGSEBQKVAVTVVDCDTGV 1491  
 DB 1507 AGESTTYPAPVPSGVSDGRHMSVOVQYNNKPNIGHGLPHGSGEKVAVTVVDDCDTGM 1566  
 QY 1492 ALRFGSVLGNCSAAOGVQSGSKSLDTGSPLLIGVVDPLPESPPVNRORVCGKRNIOY 1551  
 DB 1567 AVRREKDIIGNYSCAAQGYQTSKSLDTGSPLLIGVVDPLPESPPVNRORVCGKRNISV 1626  
 QY 1552 DSRHIDMADFLANNGVTVGCPKAKVNCDSNTCHNGGTCVNOQMDAFSCCEPLFGSGSKAQ 1611  
 DB 1627 DGKAVDAMAGFLANNGTREGCAARBPNFDCGRCCQNGGTCVNRMMNYLCECPALFPGKICEQ 1666  
 QY 1612 EMANPOHFLSSLYAMHGLSLPISQOPWYLSLMPTRQADGVLLQALTRGSTITLQREG 1671  
 DB 1687 AMPHPOHFLSGSVSWSLNITISVPWYLGIMFTRKEDSVLMEATSGSPSFLQILINN 1746  
 QY 1672 HMLSVBEGTGLQASLRLPERANDGMHQAQL---GASGGGHAI-LSFDDGQOQAE 1727  
 DB 1747 YLOREVSHGPEVDVSWLSGLRVTDGEWHLLILKRVKEDSEMKHLYTMMTLTGMDQNK 1806  
 QY 1728 GNLPRLHGLHLSNITVYGGIPGAPGVARGFRGCLQGVRSVDRPBGVSLDSPSHGSIINV 1787  
 DB 1807 ADIGMPLPGLTVRSVYVGAASEDKVSRFRGRGCMQGVBRMGCTPVNTATLMMNNALAKRV 1866  
 QY 1788 EOGCSLDPDCDSNCPNANSYSDNDWDSYSCSDPGYTGDNCTNVCDLNPCEHOSVCTRKP 1847  
 DB 1867 KDGCDVDPPCTSSPCPPNRSRCHDAMBDYSVCVCKGYLGINCVDACHNPPCENMGACVRSF 1926  
 QY 1848 SAPHGTYCECPNNTLYGYCETRIDOPCRGMWGHPTGCPNCDVSKGPDPCNTSGECH 1907  
 DB 1927 GSPGYCECBSPSHYGYPCENKDLPCPRGMWGNPVGCPGCAVSKGPDPCNTKNOQCO 1966  
 QY 1908 CKNHYPSPGSPCTCLCDYPTGSLSRVCDPBDQCCCKEGVIGRQCDRCNPPAFAYTTN 1967  
 DB 1987 CKENYTYLADOTCLPCDCCPHGSHSRCTDMACTQCAKCGVIGRQCNCRNCPNPAFYTTL 2046  
 QY 1968 GCEVNYDSCRAITAGITWMPRTFRGLPAAAPCPFGSGTAVRCHDERGMLPPLVFNCTS 2027  
 DB 2047 GCEVYINGCPAFAGITWMPQTEKGPAAVPCPGSGVGNVRHSGSGKGMPLPPLFNCTT 2106  
 QY 2028 ITFSELKFAERLORANSSGDSGSOOLALLRATQHTAGYFGSDVAVYQOLTRLLAH 2087  
 DB 2107 ISFVLDLANNMKLSRMTQVDBARALDLYALASATQHTGTLFANDVTRAYQLGHWLOH 2166  
 QY 2088 ESTORGGLSATODVHETNMLLRVGSALDTPANKRHWELIQQTEGGTAMLLQHTAAYASA 2147  
 DB 2167 ESMOQGGDLATQADQADHEHVIDHSHGSALLABATYAAWEQIORSBGGTAQILRLREGVFSN 2226  
 QY 2148 LAMQNRHTTYSPTIYTPNIVISVRLDKGNPAGAKLPREALRGSEOPDLETTVILPBS 2207

```

Db 2227 VAANRRRTYLRPFVITYANNMILAVDIFDKENFTGARVPREDTIEEPPELESSVSPAD 2286
Qy 2208 VFR-----EPPVVRPAG-----PGEAQBPPEELARQRRPPELSQGSASVITYYT 2254
Db 2287 FFRPPEKEGPIRLRPRGRTTTPOTTRPGCTERBAPISRRRRPDDAGPAAVALITYYT 2346
Qy 2255 LAGLLPHNYDPDKRSLRVPKRPPIINTPVVSIIVHDEBELPRALDKPVTVQFPLETBER 2314
Db 2347 LQQLLBERYDPRRSLRLPRPIINTPMVSTLYSEGAFLPRPLERPVLYEFALLVEER 2406
Qy 2315 TRKICVFMNHSILVSGGWSARGCEVFPNESHVSQCCHMTSPAVLMDVSRREGEIL 2374
Db 2407 TKVGVFMNHSILVSGGWSARGCELSNRKTHVACQCHTSPAVLMDISRRREGEVL 2466
Qy 2375 PLKTLTYVALGVTALAILTPFLTLRLIRSNQHGIRRLTALGAOLVFLGIGNOD 2434
Db 2467 PLKIVTYAAVSLSLALLVAFVLSLVKRLSLHSHKGLAVALLPSQLVFTVIGINOR 2526
Qy 2435 LPPACTVIALILHFLYCTFSMALBELALHLYRALTEVDVNTGPMKPYTMLGVNPAFTT 2494
Db 2527 NPFLCTVVALILHYIYMTSTFAMTLVESLHYRMLTEVRNIDTGPMKPYVVGWIGIPALYT 2586
Qy 2495 GLAVGLDPBGYNPDCCWLSIVDTLWSPAGPAPAVSMVFTYIILAAASCAAGQGB 2554
Db 2587 GLAVGLDPQGYNDPCWLSLQDTLLWSFAGPIGAVIINTVTSLSARVSCGRKHYYG 2646
Qy 2555 KKGPVGLQPSFPAVLLLSATWLLALSVNSDTLLFHYLPATCNCIGPPIFYVYVLSK 2614
Db 2647 KKGIVELRLRAFLILLILISATWLLGLLAVNRDALSTHYLPATISGIGQFPVLLFHCVLNQ 2706
Qy 2615 EVRKALK-LACSKRPDPALTTKST-LTSSYNCPSPYADG-RLYQPYGDSAGSLHSTG 2670
Db 2707 EVRKHLKGVLGGRKLLHEDSATTRATLLTRSLNCTTFGDPGPMRLTDLGESTASLDSIV 2766
Qy 2671 RS-----CKSQPSYIPFLIREESALNPGQGPGLGDSGLFLEGODQ 2713
Db 2767 RDEGIQKGVSGSLVNGSHGEPASLMPKSCXK-----PPG----- 2802
Qy 2714 HDEPDTSDSLLEDQSGSYASTHSDSEEEBEEBAAFPQEGQMDSLGPGAEPLP 2773
Db 2803 --HDSQSDSEL-DEQSSSYASHSDSDSDGVABEK-----MDPAR-----A 2845
Qy 2774 LHSYPR-DGPGPGCKAPWPDFTTAKESGNGAP-----BERLRNGDAL 2818
Db 2846 VHSYTPKDAVANHVPMGMPQSLAESDSBPCKPRLKYTVKYSVEILHREQSHGEXEP 2905
Qy 2819 SRGSGPLPGSSAOP---HKGILKKKCL---PTISEKS--SLALPLPQCTGSSRGSS 2869
Db 2906 PDBSGGAARLASQPEQRKGIKKNVYPPPLTTEQTLKRLBRKLADCSQSPSSR 2965
Qy 2870 ASBGSRGSP---PPRPPRQSLQBOQLNGVPMPLAMSIKAGTVNDESSGSE 2915
Db 2966 TSSLGSGGPPCAITVYSRGRBRGRDLNGV---AMVVRIGSAGADSDSDS 3012

```

RESULT 11

ABP81979  
ID ABP81979 standard; Protein; 3014 AA.

AC ABP81979;

DT 04-MAR-2003 (first entry)

DE Human GPCR CELSR1/Flamingo protein SEQ ID NO:444.

XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;  
 KW G protein-coupled receptor; modulator; antibody; immune-related disease;  
 KW growth-related disease; cell regeneration-related disease; AIDS; cancer;  
 KW immunological-related; cell proliferative disease; autoimmune disease;  
 KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;  
 KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;  
 KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;  
 KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;

KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;  
 KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;  
 KW ulcer.  
 OS Homo sapiens.  
 PN WO200261087-A2.  
 PD 08-AUG-2002.  
 PF 19-DEC-2001; 2001WO-US50107.  
 PR 19-DEC-2000; 2000US-257144P.  
 PA (LIFE-) LIFESPAN BIOSCIENCES INC.  
 PI Burner GC, Roush CL, Brown JP;  
 DR WPI; 2003-046718/04.  
 DR N-PSDB; AB242827.  
 PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
 PT (GPCR), useful for diagnosing and designing drugs for treating  
 PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
 PT cancer or autoimmune diseases  
 PS Disclosure; Fig 1; 523pp; English.

CC The present invention describes antigenic peptides (1) comprising:  
 CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
 CC acids. Also described: (1) an assay for the detection of a particular  
 CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
 CC and (2) an isolated antibody having high specificity and high affinity  
 CC or avidity for a particular GPCR. (1) can be used as GPCR modulators and  
 CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
 CC an antibody against a particular GPCR, and in the production of specific  
 CC antibodies. The peptides and antibodies are also useful for detecting the  
 CC presence or absence of corresponding GPCRs. The antigenic peptides for  
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
 CC treating immune-related disease, immunological-related cell proliferative  
 CC regeneration-related disease, immunological-related cell proliferative  
 CC disease, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
 CC any other disorder in which GPCRs are involved. The antibodies may be  
 CC used in immunoassays and immunodiagnosis. AB24523 to AB242869 encode  
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
 CC exemplification of the present invention.

XX Sequence 3014 AA;

Query Match 56.3%; Score 8754; DB 24; Length 3014;  
 Best Local Similarity 55.8%; Pred. No. 0;  
 Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;

```

Qy 12 TTPPPLLLILLILPPLLGQVGPCKSLGSRGSSGACAPMGWLCPSASNLMLYTSR 71
Db 57 TFRAPRELL-----DVGDDGRLAGR-RKVSAGGRPL---PLQVRLV----- 93
Qy 72 CRDAGTELTGHLVPHHDGLRVWCPSBAHPLPAPBEGCPMSCRLLGIGHLSPOGKTL 131
Db 94 ASASATLASRL-----RAKTHLP-----GCGARARLCGIGARLC--GALCF 133
Qy 132 P-----EENPCLXA-----PRLRCQS-----CKLAQAPG- 155
Db 134 PVPGGCAAGHSLAAPTLLPACRCPPRPCCRPICLPFGGSVRLTLCLALRRAGA 193
Qy 156 LRAQ-----ERSPESLGRKRRNVNTAP-----QGPFSYQAT 189

```

Db 194 VAVGALAEATAGTSPASPSPPPLPBNLPPEARAGARBARAGTSGKSLKTPMNYOVA 253  
 QY 190 VENOAGTAVASLRAIDPDEGEAGLEXTMDALPFSRGNOPFSLDPVTGAATTAEBDR 249  
 Db 254 LEFENPAGLILQIAHAYTIEGEBEVSATYMEGLPDERSGYFRIDSATGASTOSVDR 313  
 QY 250 EYKSTHVFRTVADHGMPPRSALATLITLVTDNDHPVFOOEKESLREMLLEVGYEVL 309  
 Db 314 EYKSTHVFRTVADHGMPPRSALATLITLVTDNDHPVFOOEKESLREMLLEVGYEVL 373  
 QY 310 TYRATDGAAPPANILLYRLBEGSGSPSVFEDRSGYTRGRGPRDREVSQYLTVEA 369  
 Db 374 TIRASDRSPINANIRYRLVGGAM----DVQLNBSGVSSTRVADREBAEYQDLYEA 429  
 QY 370 SPOGCPGRSTTAAVFLSVEDDNDNAPQFSEKRYVQVREDVTGAPLRYATASDRDG 429  
 Db 430 NDQGNPGLSATATYVIEVEDENDNYPOFSEONVVOVPEVGLANTAVLQOATDRDG 489  
 QY 430 SNAVVHYSIMSGNAGQFYLDAGTALDVASPLDYETTKETTLRYARADGGRPLSNVG 489  
 Db 490 QNAAHYSILSGNVAGQFYLSGLIDYINPLDEFEDVQKSLSIKAQDGRPLINSBG 549  
 QY 490 LVTQVLDNDNAPFVSTPRQATLESYPLGYLVLAHQAIDADGDNARLEYRLA---- 545  
 Db 550 VVSQVLDVNDNBPFLVSSPFOATYLENVLGYPVVHQAQVADADGENARILRYLVDTAS 609  
 QY 546 ---GVGH-----DEPTINNGTGWISVAALDREEDVPYFSGVEARHGTPLTA 592  
 Db 610 TPLGGSGSAGPKAPAPTPDPFQIHNSSGHTYCAELDREVEHYSPGVAADHSGPPMS 669  
 QY 593 SASVSVTLVDVNDNPTQPEYVVALNEDAAVGTSVTASVADDAHSVITYQITSGNT 652  
 Db 670 STSVSITVLDVNDNPEVFTQPTYEELINEDAAVSSVLTQARDRANSVITYQLTGNT 729  
 QY 653 RREBITQSGGGGLVSLALPYLKEKOVYLVATASDGRQOTAOIVVNTANTHRPVP 712  
 Db 730 RRRPALSORGGGLTLALPLDYKOBQOYVLAVTASDGRSHTAHLINVTANTHRPVP 789  
 QY 713 QSSHVTVANNEDRPAGTTVLLISATDEDTGENARITYFMEBSIPOFRIDADGATVTOAE 772  
 Db 790 QSSHVTVANNEDRPAGTTVLLISATDEDTGENARITYFMEBSIPOFRIDADGATVTOAE 849  
 QY 773 LDYEDQVSTTALITARDNGIPKXSTTYLLEILVNDVNDNAPQFLDSDYQSVYEDVPPT 832  
 Db 850 LDYEDQVSTTALITARDNGIPKXSTTYLLEILVNDVNDNAPQFLDSDYQSVYEDVPPT 909  
 QY 833 SVLQISATDRDSGLNGRVYTTROGDDGDGDIVESTSGIVTARLRENTAOVLEAY 892  
 Db 910 SILQVSATDRDSGLNGRVYTTROGDDGDGDIVESTSGIVTARLRENTAOVLEAY 969  
 QY 893 AVDKAMP--PARTPMEVTVTLVDVNDNPPVFEODEDEPVFEENSPIGLAVARTATDPBG 951  
 Db 970 AVDRSPPTPLASVETIQTIIIDINNAHPEKDELELFEENNPVGSVAKIRANDPBG 1029  
 QY 952 TNAQIMYQIVBENIDPEVQDIFSGELTALVDLYEDREPYLVTOATSAPLSVATYAV 1011  
 Db 1030 PNAQIMYQIVBENIDPEVQDIFSGELTALVDLYEDREPYLVTOATSAPLSVATYAV 1089  
 QY 1012 RLDDNDNPPVYGNBELLFNNVTVRRSSSPFGALGRVAPHPDISDLSLYFEEGNLS 1071  
 Db 1090 LTVNDNDNPPVYGNBELLFNNVTVRRSSSPFGALGRVAPHPDISDLSLYFEEGNLS 1149  
 QY 1072 LVLNASTGELKSLRDLNDRPLEAISMVLSVSDGHSVTAQCALRVTITIDEMLTHSTL 1131  
 Db 1150 LVLNASTGELKSLRDLNDRPLEAISMVLSVSDGHSVTAQCALRVTITIDEMLTHSTL 1209  
 QY 1132 RLEDMSPERFSLPLGLPTQAAVATLAPPDVHVVFNQRTDAPGHIANTLSVGP 1191  
 Db 1210 RLEDMSPERFSLPLGLPTQAAVATLAPPDVHVVFNQRTDAPGHIANTLSVGP 1268  
 QY 1192 GGGGPPPLPSBRLDRLVNLSTLTAISQVLPDPDNI CLRBCENMRCVSLARDS 1251

Db 1269 GVRG---QFPSEBDLQEOYLINRLLTITSTQVLPDPDNI CLRBCENMRCVSLARDS 1326  
 QY 1252 SAPFIASSSVLFRPIHPVGLRCRCPGFTGDYCEVEVDLCYSRPGHRCRSREGYT 1311  
 Db 1327 SAPFIASSSVLFRPIHPVGLRCRCPGFTGDYCEVEVDLCYSRPGHRCRSREGYT 1386  
 QY 1312 CLCRDGYTGEHCEVSARSGRCTPGVCKNGGTQVNLVGGFKDCPSGDFEKRYCQVTRS 1371  
 Db 1387 CECEFDPEGBHCEVDARSRCANGVCKNGGTQVNLVGGFKDCPSGDFEKRYCQVTRS 1446  
 QY 1372 FPAHSFIFRGLRQRFHHTLALSPATKRGDGLLYNGRFNRKHPVLEVOEQVLTFS 1431  
 Db 1447 FPOSFVYFRGLRQRFHHTLALSPATKRGDGLLYNGRFNRKHPVLEVOEQVLTFS 1506  
 QY 1432 AGESTTVSPVPGVSDGQNHVTLKYNPELQGTQLPGSPSOKAVATVVGCDGTGV 1491  
 Db 1507 AGETTIVAPVPSGVSDBGRHSVQVYKNGPNIHGLPLHPSGSEKNAVTVDDCOTTM 1566  
 QY 1492 ALRGSVLGNTSCAAQGTGGSKKSLDTGPELLGVBPDLSESPPVRMRQFVGCRLQV 1551  
 Db 1567 AVRFGKIDGNTSCAAQGTGGSKKSLDTGPELLGVBPDLSESPPVRMRQFVGCRLQV 1626  
 QY 1552 DSRHIDMADEIANNGTVPCCAKKVCDSNTCHNGCTCVNQWDAFSCCEPLRGFGKCAQ 1611  
 Db 1637 DGKNDVAGFIANNGTREGCARRNFCGRRQNGGTQVNMNMYLCECPILRGKNCBQ 1686  
 QY 1612 EMANPOHPLSGSLVAMHGLSLPIQOPWYLSMPFTRQADGVLLQATRGSTITQLREG 1671  
 Db 1687 AMPHROLESSESVSSKSLNIIISVPYLGIMFTRKEDSYLMARATSGSPFRQLQIAN 1746  
 QY 1672 HVMLSVEGTGLQASLRLPEGRANDGWHAQLA---GASGGPHAI--LSFDYGOQRAE 1727  
 Db 1747 YLOFESHGSPDVSVMLSGLRTVDGEWHLLILKRYKEDSEMKHLVMTLTDYGMQNK 1806  
 QY 1728 GNLGRLHGLHLSNITYGILGPRAGVYARGRGLQGRVSDTBEGVNSLDSHGESINV 1787  
 Db 1807 ADTGMLPGLTVRSVVGASSEDVSVYRGRGCGQGRMGCTPVNATVLMNNALAKRV 1866  
 QY 1788 EOGSLPDPDCNSNCPANSYCSNDMDSYS CSCDEGYGDNCTNVCDLNPCEHSGVCTRP 1847  
 Db 1867 KOGCDVDPCISSPCPNRSCHDAMBOYSVCYDQYLGINCVDAHNLPCENMAGCVASP 1926  
 QY 1848 SAPHGTCCECPNYLGYCETRIDQPCRGWGHPTGCPNCVDYSKGPDPCKNTSGECH 1907  
 Db 1927 GSPGYVCECPSHYGPCEKRLDLPCBRGMWGNVPCGPHCAVASKGPDPCNKNGCQ 1986  
 QY 1908 CKENHYRPGSPGTLCDQVYTGSLSVCPDEBQCCPKPVITRQDCRCONPFAEVTN 1967  
 Db 1967 CKENHYRPGSPGTLCDQVYTGSLSVCPDEBQCCPKPVITRQDCRCONPFAEVTN 2046  
 QY 1968 GCEVNYDCSPRAIBAGIWMPTREGLPAAAPCPKGSFGTAVRACDEHRGMLPMLFNCTS 2027  
 Db 2047 GCEVNYDCSPRAIBAGIWMPTREGLPAAAPCPKGSFGTAVRACDEHRGMLPMLFNCTS 2106  
 QY 2028 ITSEELKFAERLORNSGLDSSGSQLALLRNATQHTAGYFSGDYKAVAYQATRLIAH 2087  
 Db 2107 ISFVDLAMEKKSIRNETQVNGARALQVRLRSATQHTGTLFQNDVYATQQLGHVLOH 2166  
 QY 2088 ESTORGFLASTQVHFTENLTVRSALDTPANKRHMLQOTEGGTAMLOHAYASA 2147  
 Db 2167 ESMQGFPLAQTQADPHBEDVTHSGSALLADPATMAHBOJORSGGTQQLRLRIGFSN 2226  
 QY 2148 LAONMRHTVLSPTTIVPNIIVSVRLDKGNFAGAKLPRYEALRGBOPLDETIVLPS 2207  
 Db 2227 VARVVRTYLRPFVITVANNMLAVDIDKFNFTARVPRFOTIHBFRPRLBESSVSPAD 2286  
 QY 2208 VFR----ETPVVAPAG-----PBAQEPBELARQRRHRLBLSGEAVASYIYRT 2254  
 Db 2287 FFRPPEKEGFLAPARRTTPQTRBPGTEREAPISRRRRHDDAQFVAVALIYRT 2346  
 QY 2255 IAGLPHNYDDPKSLVYKXPIINTPVVSSVYDDEBELPRALDKPVTQVFRLLTEER 2314  
 Db 2347 LGQILPERYDPRKSLRPLRPIINTPVVSTLVSEGAFLPRFLERVLVFAALLVEER 2406



QY 302 LEVGEVUTVATDGDAPPNANILYRLLEGSGGSEVFEIDPNSGVIRTEGPDVREVE 361  
Db 366 LEVGEVUTVATDGDAPPNANILYRLLEGSGGSEVFEIDPNSGVIRTEGPDVREVE 421  
QY 362 SYOLIVASDQGRDPRSTTAAVFLSYEDDNDNAPQSEKRYVYQVVEDTPGAPVRY 421  
Db 422 EYOLLVANDQGRNRPISATRTVIEVEDENDNPQSEKRYVYQVVEDTPGAPVRY 481  
QY 422 TASDRKSNVWHSINSNGARQGYLDAQDGLDVVSPLDYETTEKTYLVRADQGR 481  
Db 482 QATBDQONQNAIHYSILSGNVAQGYLHLSGLDVLNPLDFEDVQKYSISIKQDGR 541  
QY 482 PPLSVSGUYYVQVLDINDNAPIFSTPQATVLSVPLGYLVHQAIDADQDNARLE 541  
Db 542 PPLINSSGVSVQVLDVNDNEBIFVSSPQATVLENVLPYVPHIQAVIDSGENALH 601  
QY 542 YRLA-----GVGH-----DPEFTINGGWTISVAELDREBVDYSGVEARD 584  
Db 602 YRLVDYATFLGGSSAGKRNAPPTDPFQOLHNSGMLTVOCHELDREBVDYSGVEARD 661  
QY 585 HGTPALTASASVTVLDVNDNPPPTQPEYTVRLNEDAAVGTSVTVSAVDRAHAYIT 644  
Db 662 HGSPMSSTSVSLTVLDVNDNDPVFTQPYELRLNEDAAVGSSTVTLQADRDANSYIT 721  
QY 645 YQITSQNRNRPSTISQSGGVLALPLDYKLEQYVLTANTASDGTRODPAQVANTD 704  
Db 722 YQLTQNRNRPALSSQSGGVLALPLDYKLEQYVLTANTASDGTRODPAQVANTD 781  
QY 705 ANTRRPVQSHYTVNVNEDPACTTVLISATBDNENARITYFEMDSIPORINDAT 764  
Db 782 ANTRRPVQSHYTVNVNEDPACTTVLISATBDNENARITYFEMDSIPORINDAT 841  
QY 765 GAVTTOALDEDOVSYTATITARDNGIPKSDTYLLEILVNDVNDNAPQELRDSYQGSV 824  
Db 842 GMYTMEHLDYENQVAYTLTMAQONGIPQKSDTYLLEILVNDVNDNAPQELRDSYQGSV 901  
QY 825 YEDVPFVSUQIATDSDGLNGRVFTTQSGDGDGDPVYESTSGIVTLRLDRNV 884  
Db 902 FEDAPSTSIQVSAITDSDGPNGLTYTQSGDGDGDPVYESTSGIVTLRLDRNV 961  
QY 885 AQVYLAAYVDKAMP-PARTMEVTVTVLDVNDNPPVEODEBPDVFEVENSPIGLAVARV 943  
Db 962 AVYLMALAVDRGSTPLASAVEIOVTLINDNAPMEKDELEPVEKNPVGSVAKI 1021  
QY 944 TATPDDEGTNAQIMYQIVEGNIPEVQOLDFSGELTALVDLYEDREPYLVIGATSAPL 1003  
Db 1022 RANPDDEGTNAQIMYQIVEGNIPEVQOLDFSGELTALVDLYEDREPYLVIGATSAPL 1081  
QY 1004 VSRATVHRLDRNDNPPVLDGNFELFNMYVTNNSSTFGAIGRVPHNDPDISDLYS 1063  
Db 1082 VSRATVHRLDRNDNPPVLDGNFELFNMYVTNNSSTFGAIGRVPHNDPDISDLYS 1141  
QY 1064 FERGENELSLVNLNASTGELKLSRALDNNRPLEAIVSVVS--DGVASVTAOCALRTYITD 1122  
Db 1142 FERGENELSLVNLNASTGELKLSRALDNNRPLEAIVSVVS--DGVASVTAOCALRTYITD 1201  
QY 1123 EMLTHSITLRLDMSPERFISPLGLTQAVAAATLATPDDHVVFVNVQRTDABGHILN 1182  
Db 1202 EMLTHSITLRLDMSPERFISPLGLTQAVAAATLATPDDHVVFVNVQRTDABGHILN 1260  
QY 1183 VLSLVGQPPGPGGPRPLPSDLOQRULYNSLTLTAISQVLPDDNICARECENYMR 1242  
Db 1261 VLSLVGQPPGPGGPRPLPSDLOQRULYNSLTLTAISQVLPDDNICARECENYMR 1318  
QY 1243 CVSVLRPSSAPFLASSSVLFRPIHPVGLRCRCPGFTGDCETEVLDCTSRCPGPHGR 1302  
Db 1319 CVSVLRPSSAPFLASSSVLFRPIHPVGLRCRCPGFTGDCETEVLDCTSRCPGPHGR 1378  
QY 1303 CRSRBGGYTCLCRDGTGEHCEVSARSRCCTPGVCNKGCTCVNLVGGFKDCPSGDEFK 1362  
Db 1379 CRSRBGGYTCLCRDGTGEHCEVSARSRCCTPGVCNKGCTCVNLVGGFKDCPSGDEFK 1438  
QY 1363 PYCQVTTSPFAHSITFRGLRQRHFTLALSPATKERDGLLVNGRNMGHDVVALSVI 1422

Db 1439 PYCQVTTSPFAHSITFRGLRQRHFTLALSPATKERDGLLVNGRNMGHDVVALSVI 1498  
QY 1423 QEOVQUTFS--AGSSTTVSPVPGVSGDOMQVOLKTYNK-----PLGQT 1468  
Db 1499 DEQVQUTFSAGAGETITVAPKPSGVSDBGKMSVQVQYTNKRMAPPPLPGPQPNIGHL 1558  
QY 1469 GLPQSPBOKVAVTVVDCDCTVALRPGSVLGVNSCAAGSTOGGSKSLDTGPLLGGV 1528  
Db 1559 GLPHGPGERKMAVTVVDCDCTVALRPGSVLGVNSCAAGSTOGGSKSLDTGPLLGGV 1618  
QY 1529 PDLBSPVPMRQPVGCMRLQVDSRIIDMADPLANNGYTPGCPAKONVCDNSTCHNGCT 1588  
Db 1619 PDLBSPVPMRQPVGCMRLQVDSRIIDMADPLANNGYTPGCPAKONVCDNSTCHNGCT 1678  
QY 1589 CVNOMDAPSCPCPFGGSCCAQOMAPQHFLSGLVAMGLSLPISQWYTLSPMPTRO 1648  
Db 1679 CVNOMDAPSCPCPFGGSCCAQOMAPQHFLSGLVAMGLSLPISQWYTLSPMPTRO 1738  
QY 1649 ADGVLLQAITRGSSTITLQREBHVMLSVBGTGLQASLRLBPGRANGDMHNAQLA-- 1706  
Db 1739 EDSVLMATSGGPTSPRLQILNNYVLOFEVSHGSPDVSVMLSGLRYTDMGWHLLIELON 1798  
QY 1707 -GASGGGHAL-LSFDYGOQARAGNLAPRLHGLHLSITTYGGLPGRPGYARQGRCLQG 1764  
Db 1799 VKDESEKHLVYMTLTDGMDQNRADIGMLPGLTVSVVVGASBDRKYVRGRGCMOG 1858  
QY 1765 VRVSDTPGEGVNSLDPSSHSINVEQCSLPDPCSNCPANSVCSNMDWSYSGCDPGY 1824  
Db 1859 VRMGITTYNATLMMANALKVRKXQDGDVDPCTSSCPNSRCHDMEDYSCYCDNGYL 1918  
QY 1825 GDNCTNVCDLNPCEHOSVCTRKPSAPHGTYCEBPNTLAGYCESTRIDQCPRGMGHPTC 1884  
Db 1919 GINCVDACHLNPCEHMAVCRSPGSPGCGSHYGYCEKMLDLPBPGMGHPTC 1978  
QY 1885 GPCKVDYSKFPDPCNKTSGECHKENHRRPQSPPTCLDCYPTGSLSVCPBEDQCP 1944  
Db 1979 GPCKVDYSKFPDPCNKTSGECHKENHRRPQSPPTCLDCYPTGSLSVCPBEDQCP 2038  
QY 1945 CKPGVIGRQCDNCPFAEVTNGCEVNYDSCPRATEAGIMPTRETRGLPAAACPPKGSF 2004  
Db 2039 CKPGVIGRQCDNCPFAEVTNGCEVNYDSCPRATEAGIMPTRETRGLPAAACPPKGSF 2098  
QY 2005 GTAVRHCHERGMPLPMLFNCTISITPSBELKGFARLQRNESGLDSGSQOLALLRNATQ 2064  
Db 2099 GTAVRHCHERGMPLPMLFNCTISITPSBELKGFARLQRNESGLDSGSQOLALLRNATQ 2158  
QY 2065 HTAGYSGDYKVAVOLATRLAHESQORGLSATQDVHFTENLRLAGSALLDTPANRGM 2124  
Db 2159 HTAGYSGDYKVAVOLATRLAHESQORGLSATQDVHFTENLRLAGSALLDTPANRGM 2218  
QY 2125 ELIQTEGTAMVLOHYEAVASALQNMRTYLSPTITVPTNIVISVVRADKGFAGAKL 2184  
Db 2219 ELIQTEGTAMVLOHYEAVASALQNMRTYLSPTITVPTNIVISVVRADKGFAGAKL 2278  
QY 2219 EOTORSRSGRQOLRLRLBGFSPVNAVARNRTYLRPIYIVANVLAVIDDKRFTCARV 2278  
Db 2279 EOTORSRSGRQOLRLRLBGFSPVNAVARNRTYLRPIYIVANVLAVIDDKRFTCARV 2358  
QY 2185 PRYBALKGEQPDLETTVILPESVFR---ETPPVVRPAG-----DGAQEPBELA 2231  
Db 2279 PRYBALKGEQPDLETTVILPESVFR---ETPPVVRPAG-----DGAQEPBELA 2358  
QY 2232 RQRRHPELSQGRVAVSVIYRTLAGLPHNVDQDRSLRVPKRPITNTEVVSISVUDE 2291  
Db 2339 RQRRHPELSQGRVAVSVIYRTLAGLPHNVDQDRSLRVPKRPITNTEVVSISVUDE 2398  
QY 2292 ELPLRALKRQVTVQFRLLETERTKDLCVFMNHSILVSGTGSARQCEVAFNESHVSC 2351  
Db 2339 ELPLRALKRQVTVQFRLLETERTKDLCVFMNHSILVSGTGSARQCEVAFNESHVSC 2418  
QY 2352 QCNHMTSPALVNVSRRENGELPLKTLTVVALGVTLAALLPFPPLTLRLRSNKHGI 2411  
Db 2459 QCNHMTSPALVNVSRRENGELPLKTLTVVALGVTLAALLPFPPLTLRLRSNKHGI 2518  
QY 2412 RRLNLTALGAQVFLGIGNQADLPACTYVIALPLFLVCTGSMALLBHALYRALTEV 2471  
Db 2518 RRLNLTALGAQVFLGIGNQADLPACTYVIALPLFLVCTGSMALLBHALYRALTEV 2578

```

Db 2519 HKHIAVALFLSOLVFIQINOTENPLCTVVALIHLIYNSTFAWTLVESIHYRMILETV 2578
Qy 2472 RDVNTGMRFYMYLGMVGPATITGLAVGLDPEGGNDFPCMLSTYDPLTMSFAGPAFANV 2531
Db 2579 RNIIDTGMRYTYVGMCIPIAVTGLAVGLDPOGIGNDFCMLSQDPLTMSFAGPAFANV 2638
Qy 2532 SMSVFLYIILAAASCAAROGFEKKGPSVGLQSPFAVLLLSATWLLALSVNSDTLLFFH 2591
Db 2639 IIMVTSVLSAKVSCQKHYYGKGIIVSLRTAFILLLLISATWLLGLAVNRDLNLFH 2698
Qy 2592 YLPAITCICIGPPFLISYVLSKVRKALK-LACSRPSPDPALTKST-LTSSYNCPSF 2649
Db 2699 YLPAISFGIQQPFVLLPHCVLNOBRKHLKGVLGGRKLMLEDSTYTRATILLTSLNCMT 2758
Qy 2650 YADG--RLYQPYGDSAGSLHSTSR-----GKQPSYIPLLREBSAL 2690
Db 2759 FGGPDMRLTDLGESHSTLSISYRDEGIQKLGVSSGLVRSKSGPDPASLMPRCKD---- 2814
Qy 2691 NPGQPPGLGDPGSLFLEGDDQHDPPTSDSDLSLEDDSGSYASTHSSDSSEEEEEE 2750
Db 2815 ----PG-----HDSDSSELST-DEQSSSYASHSSDSDEDGVGAE 2851
Qy 2751 EBAAPFEGQWDSLILGAEARLPLHSTPK-DGGGPGKAWPDPGTTAKSSGNAP-- 2807
Db 2852 EK-----WDPARG-----AVHSTPKDAVANHVPAQMPQSLAESDSEDPGKPRL 2897
Qy 2808 -----BERLRENGDALSRBSGLAPLPSSAQF--HKGILKKCL----PTISE 2849
Db 2898 KVEIKVSVELHREHGSHREYPPDQESGAARLASQPEQOSILKNKTYPPPLTLTE 2957
Qy 2850 KS--SLIRLPLEOCTGSSRSSASBSGSGAP---PPRPPROSLQOLNGVMPITMSIK 2903
Db 2958 QTLKGRIRKTLADCEQPTSSRTSLSGGPDCAITVKSPEGRGRDLNGV---AMNVR 3014
Qy 2904 AGTVDDSSGSE 2915
Db 3015 TGSADGSDSE 3026

RESULT 13
AAW27161
ID AAW27161 standard; Protein; 2707 AA.
XX
AC AAW27161;
XX
DT 08-DEC-1997 (first entry)
XX
DE Mouse receptor ME2.
XX
KW Chemical screening; epidermal growth factor; EGF; repeat; cadherin;
KW neuronal cell division; differentiation; cell physiology; trauma;
KW neurodegenerative disease; adenylate cyclase; cAMP.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT 1..132 /label= CD1
FT /note= "Cadherin-like repeat"
FT 133..239 /label= CD2
FT /note= "Cadherin-like repeat"
FT 240..349 /label= CD3
FT /note= "Cadherin-like repeat"
FT 350..466 /label= CD4
FT /note= "Cadherin-like repeat"
FT 467..568 /label= CD5
FT /note= "Cadherin-like repeat"
FT 569..671 /label= CD6
FT Region

```

```

FT /note= "Cadherin-like repeat"
FT 672..778 /label= CD7
FT /note= "Cadherin-like repeat"
FT 779..883 /label= CD8
FT /note= "Cadherin-like repeat"
FT 884..989 /label= CD9
FT /note= "Cadherin-like repeat"
FT 1128..1145 /label= Div EGF
FT /note= "Divergent epidermal growth factor-like repeat"
FT 1146..1178 /label= EGF1
FT /note= "Epidermal growth factor-like repeat"
FT 1186..1219 /label= EGF2
FT /note= "Epidermal growth factor-like repeat"
FT 1432..1463 /label= EGF3
FT /note= "Epidermal growth factor-like repeat"
FT 1655..1686 /label= EGF4
FT /note= "Epidermal growth factor-like repeat"
FT 1690..1724 /label= EGF5
FT /note= "Epidermal growth factor-like repeat"
FT 1773..1809 /label= DIVERGED_EGF
FT /note= "Divergent epidermal growth factor-like repeat"
FT 2244..2270 /label= TM1
FT /note= "Transmembrane portion"
FT 2282..2295 /label= TM2
FT /note= "Transmembrane portion"
FT 2309..2333 /label= TM3
FT /note= "Transmembrane portion"
FT 2346..2364 /label= TM4
FT /note= "Transmembrane portion"
FT 2389..2407 /label= TM5
FT /note= "Transmembrane portion"
FT 2432..2452 /label= TM6
FT /note= "Transmembrane portion"
FT 2463..2484 /label= TM7
FT /note= "Transmembrane portion"
XX
PN MO9707209-A2.
XX
PD 27-FEB-1997.
XX
PP 14-AUG-1996; 96WO-GB01980.
XX
PR 21-AUG-1995; 95GB-0017098.
XX
PA (UNITO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI Hadjantonakis A, Little P;
XX
DR MPI: 1997-165293/15.
XX
DR N-PSDB; AAT85320.
XX
XX Screening chemicals which interact with receptor contg. EGF-like
XX and/or cadherin-like repeats - to develop pharmaceuticals for
XX treating neurodegeneration, trauma, etc.
XX
PS Claim 10; Fig 10; 55pp; English.

```



XX A method has been developed for the screening of a chemical for  
 CC subsequent use as a pharmaceutical agent. The method involves contacting  
 CC the chemical with a receptor comprising epidermal growth factor  
 CC (EGF)-like repeats and/or cadherin-like repeats and determining whether  
 CC the chemical interacts with the receptor to form a complex. The present  
 CC sequence represents the mouse receptor ME2, which can be used in the  
 CC above method. The method is used to screen for chemicals which can cause  
 CC neural trauma, treat neurodegenerative disease, affect cellular physiology,  
 CC cyclase production, increase cAMP levels and promote neural growth.

XX Sequence 2707 AA:

Query Match 54.6%; Score 8495; DB 18; Length 2707;  
 Best Local Similarity 60.3%; Pred. No. 0;  
 Matches 1601; Conservative 386; Mismatches 569; Indels 98; Gaps 20;

192 ENQAPGTPVASLRATIDPDGEGARLEETMDALFDSNQPFLSDPTGAVTTABELDRRT 251  
 35 EKRKAGTAIETLRADPDGEGARLSYQMEALFDESNQYFLIDATGAVTTARSLDRRT 94  
 252 KSTHVRVTAODHGMFRSALATLTLYTDNDHDPVFEQEKESLRNLEFGEVLTAV 311  
 95 KOTHLAKVAVDHGSPRRSAAITLYTVSDTNDHSPVFEQSEKIRKLENGIEVLT 154  
 312 RATDGPAPNANILYRLSGSGSPSEVEIDPRSGVIRTRGPVDRBEVESYQLTVESAD 371  
 155 RATDGPAPNANMRVYRLLEGAGG---VPEIDRSGSVTRAVVDREAAEYQLVLEAND 210  
 372 QGRDPGPRSTTAAPLVSDVDNDNAQSEKRYVQVREDYTPGAVLAVTASDPRKSN 431  
 211 QGRNPEPLSASATVHLVDENDNPQFSEKRYVQVREDYTPGAVLAVTASDPRKSN 270  
 432 AVVHSISGNGARQGYLDATGALDVSPFLYETKEKTLRVARQDGRPPLSNVGLV 491  
 271 AAIHISVSGNKGQFYLSLGSGLVNPDPFARITETTLIKADQGRPPLINSGLV 330  
 492 TVQVLDINDNAPLTVSTPQATVLESVPVLYLVHQAIDADAGNARLEVLAQVGH-- 549  
 331 SVQVLDVNDNAPLTVSSPQAAVLEVPVGHSHVLIQAVDADAGENARLYQLVDTASTI 390  
 550 -----DPRPTINGCMISVAALDRREVDYTPFSGVEARDDHGTPLATASA 594  
 391 VCGSSVDSNPASADPPFQIHNSSGMITVCAELREVEHYSPGEAVDHSPAMSSSA 450  
 595 SVSVTVLDVNDNPPFTQPEYTVRLNEDAAVGSVTVYSAVDRDASHVTVTQITSGNTRN 654  
 451 SVSITVLDVNDNPPFTQPEYTVRLNEDAAVGSVTVLPAARDANSVTQITSGNTRN 510  
 655 RPSITSQSGGLVSLALPLDYLERQYVLAVTASDGTRODTAQIVNVTANTHRRPVQS 714  
 511 RPLSSQSGGLITLALPLDYKEROYVLAVTASDGTSHTAQVFINVTANTHRRPVQS 570  
 715 SHYTVNVDNRAGTIVTLISATDEPTGEMARITPMEDSIQFRIDADTGAVTTQAEID 774  
 571 SHYTVSVSDRVPVGSIAITISATDEPTGEMARITVYLAEDPVPQFRIDPTGTGITYTBEID 630  
 775 YEDQVSYTLAITARNDGIPQKSDTTYLTILVNDVNDNAQFRLDSYQGSVYEDVPEPTSV 834  
 631 YEDQAAVTLAITARNDGIPQKSDTTYLTILVNDVNDNAQFRLDSYQGSVYEDVPEPTSV 690  
 835 LQISATDRDGLNGRVFTYFGGDDGDPFVETSGVIRTLRRLDRENVQYVLRAYAV 894  
 691 LQVSTDRDGLNGRVFTYFGGDDGDPFVETSGVIRTLRRLDRENVQYVLRAYAV 850  
 895 DKGMP-PARTPMEVTVTVLVNDNPPVEFODEPVDVVEENSPICGLAVAVTATDDEGN 953  
 751 DKGSPPLASAGVQVSVLDINDNPPVEFODEPVDVVEENSPICGLAVAVTATDDEGN 810  
 954 AQIMTVIYEGNPIYEVQQLDIFSGELTAVDLDYEDRPEVLYQATSAFLVSRATVHRL 1013  
 811 AQITVQIYEGNPIYEVQQLDIFSGELTAVDLDYEDRPEVLYQATSAFLVSRATVHRL 870

1014 LDRNDNPVLAGNFELFNANNVYNNSSSPFGALIGRVPADHPDISLTYSPFERGNEISLV 1073  
 871 LDQNNPPELPDQFQLFNANNVYNNSSSPFSGVIGIRPAHDDPLDSLNTVTFIAGNEISLV 930  
 1074 LMANSTGELKSRALDNNRPLEBAIVSVSGVSHVTAQALRVTTITDEMTHTSITRL 1133  
 931 LMDPATGELISRLDNNRPLEBAIVSVSGVSHVTAQALRVTTITDEMTHTSITRL 990  
 1134 EDMSEERFLSPILGLFQAVAAATLPPDHVVVNVVQDTPAAGHIIINLSVQPPGP 1193  
 991 EMMSEERFLSPILGLFQAVAAATLPPDHVVVNVVQDTPAAGHIIINLSVQPPGP 1049  
 1194 GGGPPELPSEDLQERLYNRSILTAISAQVLPEDDNLCLREPCENYRCSVSLRFDSSA 1253  
 1050 RG-RFPSEDLQERLYNRSILTAISAQVLPEDDNLCLREPCENYRCSVSLRFDSSA 1107  
 1254 PFIASSVLPFPIHPVGGILRCRCPGFTGDCETREVDLCYRPPCGPHGRCSREGYTCL 1313  
 1108 PFISSVLPFPIHPVGGILRCRCPGFTGDCETREVDLCYRPPCGPHGRCSREGYTCE 1167  
 1314 CRDGYTGHCVSARSGRCTPGCKNGGTCVNLVVGPKDCDPSGDPFEPKPCOYVTSFP 1373  
 1168 CFEDPTGSHCOVNVRSGRKASGVCKNGGTCVNLVVGPKDCDPSGDPFEPKPCOYVTSFP 1227  
 1374 AHSPTTFGRGLRORFHTLALSPATKERDGLLVNGRFNEKADFYALLEVYIQVQVOLTFSAG 1433  
 1228 POSFTVTFGRGLRORFHTLALSPATKERDGLLVNGRFNEKADFYALLEVYIQVQVOLTFSAG 1287  
 1434 ESTTVSPFVPGVYSDGQMTVOLKTYNKPILGOTGLPQGSBQKAVAVTYDCCDTVAL 1493  
 1288 ETTTVTVTVQVGVGSDGMHSLVQYNNKPNIGHGLPHGSGSKRVAVVTVDDCAVAV 1347  
 1494 RFGSVLAGNSCAAGTQSGSKSLDLTGPIILGGVPLPESFPVRMPOFVQCMENLOVDS 1553  
 1348 HFGSVLAGNSCAAGTQSGSKSLDLTGPIILGGVPLPESFPVRMPOFVQCMENLOVDS 1407  
 1554 RHIDMAPIANNGTVPGCPAKKAVCDNSNTCHNGGTCVNMQMAFSCPCPLGFGKSCAQEM 1613  
 1408 RIVDMAPIANNGTVPGCPAKKAVCDNSNTCHNGGTCVNMQMAFSCPCPLGFGKSCAQEM 1467  
 1614 ANPQHFGLSSLVAMHGLSLPISQPMYLSMERTTQADGVTLQAITRGSTITLQRESHV 1673  
 1468 PHPQFRTGSESVVMSDDITTSVPMYGLMERTKEBGMVMEATVAGTSSRLHLQITANSYI 1527  
 1674 MLSEVGTGLQASSLRBERGRANDGMHQAOLALASGSGPH-----HLSFDYQGRAGE 1728  
 1528 RFEVSYGSDVANSQLSKSRITTDGGMHLLIEL-RSAKRGKDIKYLAVMTLDYGHDOSTV 1586  
 1729 NLGPRHLGHLNITTVGGIPGPAGVARGFRGLQGVRSPTPEGVNSLDPHSGESINVE 1788  
 1587 QIQNGQLPGLKRTIVIGGVTEADKYVSRHGRFGCQGVARMGSSNTIATLNNMDLAKRYK 1646  
 1789 QGCSLPDPCDSNPPCANASYCNDMDYS CSCDPQYTGDNCTNVCDLNPCHESQSVCTRKPS 1848  
 1647 DGCVDPEPCASSPCPPHRCPTDWDYSICIDRGYLRKKCVADCLNPPCKHVGSLCALPN 1706  
 1849 APHGTYCECPNMYAGPYCETRIIDQPCPRGMWGHFTGCGCNCDVKGPRDPCNKTSGBCHC 1908  
 1707 TPRGYSCGCGHGGQICENKVDLPCEPRGMWGNCAVPVYLSAKALIPYAT----- 1758  
 1909 KENHYRPGSPYCLLDCYR-----TGSLSRVCDPEDEQCPCPKGVYIGRQCRCRD 1958  
 1759 -----RPMASARRITTSPPQPIVAFVTVSPRSHSRACMDOTGCAKPGVYIGRQCRCRD 1813  
 1959 NPFAVTTNGCVNVYDSCPRALIEGIMWPRTRFBLPAAAPCPKKSFGTAVAHCHDEHGM 2018  
 1814 NPFAVTVSLGCEVLYNCGCPRAFEAGIMWPMKFTQPAVYICPKKS YGANA VYHCGEKGWL 1873  
 2019 PPNLFCNCTISFSSILKFAERLQENESGLDGSRCQALALLRNATORTAGFGSGDVAVY 2078  
 1874 PPELFCNCTISGFDVLKALNKTANNETRMQGNBSRLAKALRNATQGNSTLFGNDVATAY 1933

QY 2079 QLATRLAHSTORGFELSATODVHTENTLRYGSAALDTPANKRHWELIQOTEGTAMLL 2138  
 DB 1934 QLATRLAHSTORGFELSATODVHTENTLRYGSAALDTPANKRHWELIQOTEGTAMLL 1993  
 QY 2139 QHTBAVASALQNMRTHTLSPPTIVTENVISVVRDKGNFAGAKLPRYEALRGBOPLD 2198  
 DB 1994 RHEFAVSNVARNVKETYLARPFIIVTANMILAVDIPDKINFTGAQVPRFEDIDELPREL 2053  
 QY 2199 ETTVILPESVP-----RTPPVNR-----PAGEBAQEPBELARQRHRELQGER 2245  
 DB 2054 ESSVSPADTKEPEKKEGVVRLTNRRTPPLTAQEPERERTSSRRRRHDEBQQA 2113  
 QY 2246 VASVITRTLAGLPHNYDDPKSLRVKRPINTIPVSIYVDEDELPALADKPVTVQ 2305  
 DB 2114 VALVITRTGQLPHNYDDPKSLRVKRPINTIPVSIYVDEDELPALADKPVTVQ 2173  
 QY 2306 FLLEETERTKPICVNWNHSLVSGTGSAGCEVFNESHVSCQCNMTSPAVLMDV 2365  
 DB 2174 FSLLEETERTKPICVNWNHSLVSGTGSAGCEVFNESHVSCQCNMTSPAVLMDV 2233  
 QY 2366 SRENGEELLPLKTLTVAGVTLAALLTFEFLTLRLASNGHGRNLTALGLAQV 2425  
 DB 2234 SRENGEELLPLKTLTVAGVTLAALLTFEFLTLRLASNGHGRNLTALGLAQV 2293  
 QY 2426 FLGINQADLPFACVTAIHLHFLYCTFSMALLEALHLTRALTEVDVNTGEMRYML 2485  
 DB 2294 FMVGINQADLPFACVTAIHLHFLYCTFSMALLEALHLTRALTEVDVNTGEMRYML 2353  
 QY 2486 GWCVPVAFITGLAVGLDDEGVGNPFCMLSTYDTLLNSFAGVPAFAVMSVFTIILARRS 2545  
 DB 2354 GWCVPVAFITGLAVGLDDEGVGNPFCMLSTYDTLLNSFAGVPAFAVMSVFTIILARRS 2413  
 QY 2546 CANOQGFKEKGVSGLOPFAVILLASATWLLASVNSDTLLFHTLPTACNCTGCPFL 2605  
 DB 2414 CANOQGFKEKGVSGLOPFAVILLASATWLLASVNSDTLLFHTLPTACNCTGCPFL 2473  
 QY 2606 FLGYVVLISKERVALK-LACSRKSPDPALTYKST-LTSSYCNPPYADRLTYQPY--GD 2661  
 DB 2474 FLGYVVLISKERVALK-LACSRKSPDPALTYKST-LTSSYCNPPYADRLTYQPY--GD 2533  
 QY 2662 SAGSLHSTSGSGSQSYTYFLLRBSALNPGGPGQLDPSGLFL-EGDQDQHDPTMS 2720  
 DB 2534 SAGSLHSTSGSGSQSYTYFLLRBSALNPGGPGQLDPSGLFL-EGDQDQHDPTMS 2583  
 QY 2721 DSDLSIEDDQSGSYASTHSDSEBEEBEEBAAPFEGQMDSLPGCARLRLHSTPK- 2779  
 DB 2584 DSDLSIEDDQSGSYASTHSDSEBEEBEEBAAPFEGQMDSLPGCARLRLHSTPK- 2628  
 QY 2780 DGGPGPGKAPWPGD 2793  
 DB 2629 DALANHVPAWPGD 2642  
 RESULT 14  
 ABP53570 standard; protein; 4115 AA.  
 AC ABP53570:  
 XX 17-DEC-2002 (first entry)  
 DE Human NOV2 protein SEQ ID NO.4.  
 KM Human; NOVX: cytosolic; neuroprotective; anticonvulsant; cardiovascular;  
 KM cerebroprotective; nootropic; antidiabetic; antiinflammatory; fungicide;  
 KM antineuritic; anticholinergic; immunosuppressive; antiallergic; virucide;  
 KM antidiabetic; antibacterial; protozoacide; antihelminthic; gene therapy;  
 KM cancer; leukaemia; lymphoma; melanoma; neurological disorder; epilepsy;  
 KM stroke; ischaemic cerebrovascular disease; Alzheimer's disease; allergy;  
 KM Pick's disease; vesicular transport disease; cystic fibrosis; goitre;  
 KM diabetes mellitus; Graves' disease; gastrointestinal disorder; vaccine;  
 KM ulcerative colitis; gastric disorder; duodenal disorder; infection;  
 KM autoimmune disease; allergic reaction; autoimmune haemolytic anaemia;

KW rheumatoid arthritis; chromosome 3q21.3-4.  
 XX Homo sapiens.  
 OS WO200262999-A2.  
 XX 15-AUG-2002.  
 PD 31-DEC-2001; 2001WO-US49976.  
 PF 29-DEC-2000; 2000US-258928P.  
 PR 02-JAN-2001; 2001US-259415P.  
 PR 04-JAN-2001; 2001US-259785P.  
 PR 20-FEB-2001; 2001US-269814P.  
 PR 09-MAR-2001; 2001US-279863P.  
 PR 29-MAR-2001; 2001US-279832P.  
 PR 29-MAR-2001; 2001US-279833P.  
 PR 13-APR-2001; 2001US-283889P.  
 PR 18-APR-2001; 2001US-284447P.  
 PR 25-APR-2001; 2001US-286683P.  
 PR 29-MAY-2001; 2001US-294080P.  
 PR 16-AUG-2001; 2001US-312815P.  
 PR 17-AUG-2001; 2001US-313325P.  
 PR 17-SEP-2001; 2001US-322699P.  
 PR 26-NOV-2001; 2001US-333350P.  
 XX (CUBA-) CUBAGEN CORP.  
 XX Spytek KA, Li L, Wolenc AR, Vernet CM, Eison A, Liu X;  
 PI Malpankar U, Shinkens RA, Tchervet VT, Spaderna SK, Gorman L;  
 PI Kekuda R, Paturajan M, Gusev V, Gangoli EA, Guo X, Shenoy S;  
 PI Raselli L, Casman SJ, Boldog F, Burgess CE, Edinger S;  
 PI Elleman K, Gunther E, Smithson G, Miller I, MacDougall JR;  
 XX WPI; 2002-732706/79.  
 DR N-PSDB; ABQ82327.  
 XX New NOVX polypeptides and polynucleotides useful for treating  
 PT NOVX-associated disorders, such as cancers, neurological disorders,  
 PT disorders of vesicular transport, gastrointestinal disorders, and  
 PT autoimmune diseases -  
 PS Claim 1, Page 17-18; 444p; English.  
 XX The present invention describes novel human proteins designated NOVX,  
 CC where X is 1 to 20 e.g. NOV1. NOVX sequences can have neuroprotective,  
 CC cytosolic, anticonvulsant, cerebroprotective, nootropic, cardiovascular,  
 CC antidiabetic, antiinflammatory, antineuritic, antiallergic, fungicide,  
 CC immunosuppressive, antihelminthic activities, and can be used in gene  
 CC therapy. The NOVX proteins, nucleotides or antibodies can be used in the  
 CC manufacture of a medicament for treating a syndrome associated with a  
 CC human disease selected from NOVX-associated disorder, such as cancers  
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,  
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke,  
 CC ischaemic cerebrovascular disease, Alzheimer's disease or Pick's  
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,  
 CC diabetes mellitus, Graves' disease, or goitre), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),  
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic  
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic  
 CC and protozoal infections. The NOVX proteins can be used as immunogens to  
 CC produce antibodies and as vaccines. The NOVX nucleotide sequences may  
 CC be used in chromosome mapping, identifying individuals from minute  
 CC biological samples (tissue typing), and in forensic identification of a  
 CC biological sample. The present sequence represents human NOV2, which is  
 CC located on chromosome 3q21.3-4.  
 XX Sequence 4115 AA;  
 SQ  
 Query Match 50.3%; Score 7816.5; DB 23; Length 4115;  
 Best Local Similarity 52.2%; Pred. No. 0;  
 Matches 1555; Conservative 425; Mismatches 802; Indels 197; Gaps 43;

QY 41 GSRGSGSGACAPMGWLCPSASNLMLYTRCDAETELTG---HLVPHDGLRWCPES 97  
 Db 193 GSRKRVGTACCCBELMWTSGKGG---BRATTSGLERAPRNCPLKASG----- 229  
 QY 98 EAHIPLPRAPEGCPWSCRLLIGIGHLSPOGKLTLPBEPHCLKAPRLKQSG---CK-LAQ 153  
 Db 240 -----SGPELDSAPRTARPAASG---APRESRTAPBPAF---KRNRSGLPFCRFLPQR 289  
 QY 154 PG-----LRAGEBSPESLGGRRKNVNTAPQPOPSPAYQVPEBQAGTVAASLRAT 206  
 Db 290 PGRPPPLPAPRPAKKTTSANRABFRBANRHPQPOYNQVTLVPEBAAGTVLRVAQ 349  
 QY 207 DPDEGEAGLEYTMDALFDSRSNOFSLDPVTGAVTAAEELDRBETKSTHYFRVTAODHM 266  
 Db 350 DPDAEGAGRLVYSALALMNSRSLBESIDPOSGLIRRAALDREMSERHLYRTAODHS 409  
 QY 267 PRSALATLTLYTDTNDHPVBEQOEYKESLRENTLEVGYEVLTVRATDGDAPNNATLY 326  
 Db 410 PRLSATTMVAVTADRDHSPVEQOAYRETLREMBEGYPIILQTRATDGDAPNNATLY 469  
 QY 327 RL- -EGSGSPSEVPEIDRSGVIRTRGVDRBEVSVOLTVBASDQGRDPGRSTTAA 384  
 Db 470 RYVGPAPARAAAFAEIDPRSGLISTSGVDRBHMSYELVVEASDQGRDPGRSTTAA 529  
 QY 385 VFLSVEDNDNAPQPSBKRYVVOVREDVTEGAPVLAVTASDRDGSNAVHVSIGMGNAR 444  
 Db 530 VHTLVTDENDNAPQPSBKRYVVOVREDVTEGAPVLAVTASDRDGSNAVHVSIGMGNAR 589  
 QY 445 GQYVLAQAGALVNSPLDYETTEKYTLVRAODGGRPPLSNVSGLVTVQVLDINDAPI 504  
 Db 590 GHFAISLSTGEIOVVAPLDPEAREEYALRIADAGPPLSNNTGLASIQVVDINHPI 649  
 QY 505 FVSTPQATVLSVPLGYLVLAQVADADAGNARLEVRLAGVGHDPPTTNGCTMISY 564  
 Db 650 FVSTPQATVLSVPLGYLVLAQVADADAGNARLEVRLAGVGHDPPTTNGCTMISY 709  
 QY 565 AAELEDEVDYFSGVEARDHGTPTALVTSASVTVLVDVNDNPTTOBEYVRLNEDAA 624  
 Db 710 SGPLDRESEVHYFVGVEARDHGTPTALVTSASVTVLVDVNDNPTTOBEYVRLNEDAA 769  
 QY 625 VGTSVTVASVANDASVITYQITSGTNRNBSITQSOGGGLVSLALPLDYKEROYVLA 684  
 Db 770 VGTSVTVASVANDASVITYQITSGTNRNBSITQSOGGGLVSLALPLDYKEROYVLA 829  
 QY 685 VTASDGTRODTAQIVNVVDANRHPVFSOSHTVAVNEDRPACTVVLISADDEDTGEN 744  
 Db 830 LTHSDRALHCHCYVHNITDANTHRPVPQSAHVSVDNEDRPMGSTTVVISAADDVGEN 889  
 QY 745 ARTTPEMDSIPQFRIDADTGAATTOAELEDEVDQVSYTLATARDNGIFOKSDTTYLEIL 804  
 Db 890 ARTTPEMDSIPQFRIDADTGAATTOAELEDEVDQVSYTLATARDNGIFOKSDTTYLEIL 949  
 QY 805 VNDVNDNAPQPLDSDYQSGYTEDVPPPTVYLOISATDRDSGLNGRVYTPQGGDGDGP 864  
 Db 950 VNDVNDNAPQPLDSDYQSGYTEDVPPPTVYLOISATDRDSGLNGRVYTPQGGDGDGP 1009  
 QY 865 IVESTSGIVRTLEBRLEBNAOVVLAAYAVDKGMPARTPMETVTVLVDNNDPFPBED 924  
 Db 1010 TLEPTSGIVRTLEBRLEBNAOVVLAAYAVDKGMPARTPMETVTVLVDNNDPFPBED 1069  
 QY 925 EFDVPEVENSPIGLAVARVATATPDDEGTNAQIMYQIVEGNIPPEVOLDIFSGLTALVUL 984  
 Db 1070 EFDVPEVENSPIGLAVARVATATPDDEGTNAQIMYQIVEGNIPPEVOLDIFSGLTALVUL 1129  
 QY 985 DYEDREYVLYOATSAPLVSRAVTVRLIDRNDNPPVLCNFEILFNNTVNRSSFPGG 1044  
 Db 1130 DYEDREYVLYOATSAPLVSRAVTVRLIDRNDNPPVLCNFEILFNNTVNRSSFPGG 1189  
 QY 1045 AIGRVAHDPDLSLTYSPERGENLSIVYLNASSTGELKSRALDNRRPLAISMVLYSD 1104  
 Db 1190 IIGRVAHDPDLSLTYSPERGENLSIVYLNASSTGELKSRALDNRRPLAISMVLYSD 1249

QY 1105 GVHVTAAQALRVTTITTDMLTHSITRLBDMSPERFLSPLGLFIOVAATLATAPPDHY 1164  
 Db 1250 GLHSTVACVLANVITTEBELANSLTVRLBNMGERFLSPLGLFIOVAATLATAPPDHY 1309  
 QY 1165 VVFNVRDTPADQGHILANVLSVQPPGQGP- -PLPESDLQERLYNRSILTAISAQ 1222  
 Db 1310 VVFNVRDTPADQGHILANVLSVQPPGQGP- -PLPESDLQERLYNRSILTAISAQ 1277  
 QY 1223 RVLPPDNDICLRBPCENWRCVSLRFDSSAPFLSSSLPPIPHVGLRCCPFG 1282  
 Db 1369 RVLPPDNDICLRBPCENWRCVSLRFDSSAPFLSSSLPPIPHVGLRCCPFG 1428  
 QY 1283 DYCETEVLDICYSRPGPHRCRSREGGYCLCRDGYTGHCYVSARSAGCTGVNAGGT 1342  
 Db 1429 DYCETEVLDICYSRPGPHRCRSREGGYCLCRDGYTGHCYVSARSAGCTGVNAGGT 1488  
 QY 1343 CYNLVGFKCDPCSGD- -PEKPYCOVTTSPHSPFTTFRGLRQRFHTLALSPATKED 1401  
 Db 1489 CYNLVGFKCDPCSGD- -PEKPYCOVTTSPHSPFTTFRGLRQRFHTLALSPATKED 1461  
 QY 1402 GILLVNGRNEKHDPALAEVLOEVOVQTFPSAGESITTVSPFVPGVSDQOMHTVQLKYN 1461  
 Db 1549 GILLVNGRNEKHDPALAEVLOEVOVQTFPSAGESITTVSPFVPGVSDQOMHTVQLKYN 1608  
 QY 1462 KPLAQGTGLPQPSBQKVAVTVVDCDGTGVALRPSVLAGNYSCAAQGTGSKSLDLTG 1521  
 Db 1609 KPLAQGTGLPQPSBQKVAVTVVDCDGTGVALRPSVLAGNYSCAAQGTGSKSLDLTG 1668  
 QY 1522 PULLGVPDLPSFPVRNRPQVPCNRNTOVDSRHITDMADFIANNGTVPCCPAKKNVCSN 1581  
 Db 1669 PULLGVPDLPSFPVRNRPQVPCNRNTOVDSRHITDMADFIANNGTVPCCPAKKNVCSN 1728  
 QY 1582 TCHNGSTCNOMDASCPGPIGFGKSCAQEMANOHITGLSSLVMM- -GLSLPIQOWTL 1640  
 Db 1729 TCHNGSTCNOMDASCPGPIGFGKSCAQEMANOHITGLSSLVMM- -GLSLPIQOWTL 1788  
 QY 1641 SLMPRTQADGVLTQAITRGSSTITLQLEBHVMSV- -EGTLQASSLLEBRANDGDW 1699  
 Db 1789 SLMPRTQADGVLTQAITRGSSTITLQLEBHVMSV- -EGTLQASSLLEBRANDGDW 1847  
 QY 1700 HHAQAL- -GASGFGHAIL- -SPDYGOORAEAGNIGPLHGLHSNTYGGI- -PGPAGV 1754  
 Db 1848 HHAQAL- -GASGFGHAIL- -SPDYGOORAEAGNIGPLHGLHSNTYGGI- -PGPAGV 1907  
 QY 1755 AGFRGCCQGVAVSDTPRGVNS- -LDSHGBESINVGQSLRPPCNSNCPNAYSQSNMD 1813  
 Db 1908 AGFRGCCQGVAVSDTPRGVNS- -LDSHGBESINVGQSLRPPCNSNCPNAYSQSNMD 1965  
 QY 1814 SYSCSDGEGYGDNCTNVCIDLPCEHOSVCTRKPSAPRGYTCEPNTLYGPYCETRIDP 1873  
 Db 1966 SYSCSDGEGYGDNCTNVCIDLPCEHOSVCTRKPSAPRGYTCEPNTLYGPYCETRIDP 2025  
 QY 1874 CPGWGHGHTGCPKNCVDYSKGDPDPCNTSGEHCCKENHTRPQSPPTCLLDCYTGSL 1933  
 Db 2026 CPGWGHGHTGCPKNCVDYSKGDPDPCNTSGEHCCKENHTRPQSPPTCLLDCYTGSL 2085  
 QY 1934 RYCDPEBQCPKPCVITROCDRCNRPAYVTNCCFVNYSCPAIAGMIPRTTGL 1993  
 Db 2086 RYCDPEBQCPKPCVITROCDRCNRPAYVTNCCFVNYSCPAIAGMIPRTTGL 2145  
 QY 1994 PAAPCPKSGFQTAVRHCDERHGMPLPMLFNCTSTTSEBLKGFABRLORNSGLDSCRSQ 2053  
 Db 2146 PAAPCPKSGFQTAVRHCDERHGMPLPMLFNCTSTTSEBLKGFABRLORNSGLDSCRSQ 2205  
 QY 2054 QIALILRNATOTHTAGYFSGDYKAVAYQALATRLAHSTORGGGLASTOVHFTENLRYGS 2113  
 Db 2206 QIALILRNATOTHTAGYFSGDYKAVAYQALATRLAHSTORGGGLASTOVHFTENLRYGS 2265  
 QY 2114 ALLDTANKHMEIL- -QOTEG- -GTAMLOHYEAVASLAONMRTVYSPFTVYPNVI 2169  
 Db 2266 ALLDTANKHMEIL- -QOTEG- -GTAMLOHYEAVASLAONMRTVYSPFTVYPNVI 2325  
 QY 2170 SVRLDKGNF- -AGAKLPRYEA- -LRGQPPDLTITVILPESVFPETPPVVRPAGGEAQ 2225

```

Db 2326 SIDRMHPSPPRGARPRYHSNLFRGQAMDPHTVLLPSQSPRPSPEVLPTSSSIEN 2385
Qy 2226 EPEELARRGRHRELSQGEAVAVIYRTLAGLPPNYPDPKSLAVPKRPINTVYSI 2285
Db 2386 STTSVPPPPAPPEPPEGISIIILLVYRTLGILLPQFQERRGALLPQNPVNSPVSIV 2445
Qy 2286 SVHDEELLPRALDKPVTVQFRLLETEERTKPICVPMNSILVSGTSGMSARCEVFPNN 2345
Db 2446 AVFHGNFPLRGLESPIISLEFRLQTNANSKALCVQMDPGLACQGWARTARDELVHNN 2505
Qy 2346 ESHVSCQCNHMTSPAVLMDVSRPE--NGEILPKLTLYVALGVTLAALLTPEPFLTLRI 2403
Db 2506 GSHARCRCRCTGTFGVLMASPREERLELAVFTHVVAVALVTLAAILLSTLS 2565
Qy 2404 LRENGHGRRLTALAIQVFLGILNQ-----ADLP 2436
Db 2566 LKSNVGHANVANVAALGVALLFLGLGIRHTNNOVQDQGTCLVMTLLAQEWANGQSGE 2625
Qy 2437 FACTVAIILHFLYLCFSWALLBALHYRALTEVDVNTGPMRFYMLGMVPAFITGL 2496
Db 2626 LVCTAAVAILHVFSTFANLFGVGLHYTMQVEPRNVRGNRFTHALGMGPVAILGL 2685
Qy 2497 AVGLDPEGTGNPPFCMLSTYDTLINSFAGFVAVSMASVFLYLAAAPASCAAGQGPCK 2556
Db 2686 AVGLDPEGTGNPPFCMISVHREPLMSFAGFVAVIIVNGTMFLLARTSCSTQREAKKT 2745
Qy 2557 GPVSGLGQSPAVVLLSATVLLALSVNSPTLLFHTLPAQNCIQGPFIPLSVVYLSKEV 2616
Db 2746 SALRTLRSTFLVLLVLSASWLFGLLVANNSILMFHTLHGLCGLQGLAVLLFCVLANDA 2805
Qy 2617 RKALKLAC--SRKSPDPALTTKSTLTSSVNCSPYADGRLYQ--PYGDSAGSLHSTSRGK 2674
Db 2806 RAAMWPAICLRKAPREARPAFLGAVNNTALPESGLIRITLGAFTVSSVSAASGR 2865
Qy 2675 SQP-----SYT--PFLARESA-----LNPQGPFGIG-----DPGSLFLEGQDQ 2712
Db 2866 TDQDQSGRGSYLRDNLVHNGSAADHTDHSLOAHAGPDLDVAMFHRDGA----- 2917
Qy 2713 QHDPDTSDSLLEDQSGSYASTHSDSEEESEEEBAAPRGQWDSLIGPQABRL 2772
Db 2918 ----DBDSBSDLSEERSLSTSSSEEDNGRTGRGFORPLCAAQ-----SERL 2963
Qy 2773 PLHSTPKDGGPGPKAPWG-----DFGT-----TAKSSGNGAPBERLRE 2813
Db 2964 LTH--FKDYVGNMLISYWPALGECBAPCALQWGSERRLGLTSDKDAANNQPDAL--T 3020
Qy 2814 NGDALSRBSGLPPLPSSSAQPHKGLKKKCLPTISEKSSILRLP-----LEQCTG 2863
Db 3021 SGD-----ETSL-----GRAQRQRKGIKNR-----LQYPLVQTRGAPELSWCRA 3061
Qy 2864 SSRGSSA-----SESGRGPPPPPPROSLQEBLN 2893
Db 3062 ATIGHRAVPAASYGRYAGGCTGSLQPARVYSRRQLD 3100

```

```

KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
KW infection; cell proliferative disorder; actinic keratosis; myeloma;
KW arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotoxic;
KW neutrotective; dermatological.
OS Homo sapiens.
FH Key
FH Peptide
FT 1..31
FT /label= Signal_peptide
FT 32..3298
FT /note= "Mature human extracellular matrix and cell
FT adhesion molecule (XMAP)"
FT 329..423
FT /note= "Cadherin domain"
FT 437..535
FT /note= "Cadherin domain"
FT 549..641
FT /note= "Cadherin domain"
FT 655..746
FT /note= "Cadherin domain"
FT 760..848
FT /note= "Cadherin domain"
FT 862..952
FT /note= "Cadherin domain"
FT 965..1057
FT /note= "Cadherin domain"
FT 1071..1159
FT /note= "Cadherin domain"
FT 1178..1265
FT /note= "Cadherin domain"
FT 1378..1431
FT /note= "EGF domain"
FT 1438..1469
FT /note= "EGF domain"
FT 1478..1512
FT /note= "EGF domain"
FT 1536..1553
FT /note= "EGF domain"
FT 1542..1704
FT /note= "Transmembrane domain"
FT 1725..1756
FT /note= "Laminin G domain"
FT 1792..1920
FT /note= "Laminin G domain"
FT 1931..1962
FT /note= "EGF domain"
FT 1966..2000
FT /note= "EGF domain"
FT 2110..2137
FT /note= "EGF domain"
FT 2525..2545
FT /note= "GPCR signature"
FT 2588..2613
FT /note= "Transmembrane domain"
FT 2659..2689
FT /note= "GPCR signature"
FT 2712..2730
FT /note= "Transmembrane domain"
FT 2741..2761
FT /note= "Transmembrane domain"
PD WO200142285-A2.
PN 14-JUN-2001.
XX
XX
XX 05-DEC-2000; 2000MO-US32990.
XX 10-DEC-1999; 99US-0172852.
XX 16-DEC-1999; 99US-0172354.
XX (INCY-) INCYTE GENOMICS INC.

```

XX Yve H, Tang YT, Lai P, Burford N, Azimzai Y, Patterson C;  
PI Baughn MR, Lu DAM, Shan P, Au-Young J;  
PI WPI; 2001-361632/40.  
DR N-PSDB; AAD08065.  
XX  
XX  
PT New human extracellular matrix and cell adhesion molecules and  
PT polynucleotide sequences encoding them, useful for diagnosis,  
PT prevention, treatment of genetic, autoimmune and cell proliferative  
PT disorders -  
XX  
XX  
PS Claim 1; Page 113-119, 135pp; English.  
XX

Claim 1; Page 113-119; 135pp; English.

The present sequence is a human extracellular matrix and cell adhesion molecule (XMAP). The XMAP is used for screening a compound for effectiveness as an agonist or antagonist of XMAP. The identified agonists or antagonists are used for treating a disease or condition associated with decreased or increased expression of functional XMAP. The polynucleotides encoding XMAP are useful in somatic or germline gene therapy to correct a genetic deficiency, to express a conditionally lethal gene product and to express a protein which affords protection against intracellular parasites and also for diagnosis of disorders associated with expression of XMAP. They are also used for generating hybridisation probes useful in mapping the naturally occurring genomic sequences and to create knock in humanised animals (bigs) or transgenic animals (mice or rats) to model human diseases. Oligonucleotide or longer fragments derived from the polynucleotide sequences may be used as elements on a microarray. Antibodies which specifically bind XMAP may be used for the diagnosis of disorders associated with the expression of XMAP, or in assays to monitor patients being treated with XMAP. Diseases diagnosed, prevented or treated include genetic disorders such as adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's disease, myotonic dystrophy, sickle cell anaemia, thalassemia, autoimmune/inflammatory disorders such as acquired immune deficiency syndrome (AIDS), Addison's disease, allergies, anaemia, asthma, atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis, glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis, osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal, parasitic, protozoal and helminthic infections and cell proliferative disorders such as actinic keratosis, arteriosclerosis and cancer including breast, bladder, bone marrow, brain and uterine cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.

**SQ Sequence 3298 AA;**

Query Match	Score	DB	Length
49.9%	7763	22	3298

best local similarity 52.4%; Freq. NO. 0;  
Matches 1549; Conservative 418; Mismatches 796; Indels 194; Gaps 45

```

QY      4  GSRGGSSGACAPMGMLCPSSASNMJLTVSSCRDAGTELNG---HYVHHIDGLRWCPES  97
      193  GSRKRVGTARCCGELMAYGSKQG-----EATVTSGBARTAPRNCCIFGASG-----  239
QY      98  EAHITLPPEAPBEGCPSRLLIGIGHLSPOGKLTLPBPHCLKABRLKQS---CK-LAQA  153
      240  -----SGPELIDAPPTARTATAPASG-APRESRTAPEPAP-----KXMSRGLFRCLFPQR  289
QY      154  PG-----LRAGESPBEISGGRKRWVNTPAPOPPSYQATVPENOPAGTPASLRAL  206
      290  PGPRRPGILPAREAKKVTSSANRARRRANAHPOFPQYNTVTLVEEBAAGTALRLVAO  349
QY      207  DPDEBAGRLBYTMALPDPSRNOFSLDPVTGAVTTLAEILDRETKSTHYFRVTAODHGM  266
      350  DDPAGAGRLVYSLAAMNWSRLTEFSLDPQSGILRTAALADRESMRHYTLVAODHGS  409
QY      267  PRRSALATTLIVTNTDHDVFEEOQAEKESIRENLENGYEVLVVARTDGDAPANNATLY  326
      410  PRLSATVWAVTVADNRNHSPLYFEAQVRETILRENVESGYITLDRATDGDAPANNATLR  469
QY      327  RLL--EGSGSGSSEVFEIDPRSGVTRKTPYDREVEVSYQLTVBASQGDPEGRSTTAA  384
      470  RRVGPPAALAAAAAFAEIDPRSGLLSTGSRVDRREMEVEYELTVBASQGDPEGRSAYTA  529

```

QY	385	VFLSVEEDNNAPQFSSEKRYVVOQREBVTGAPLRYATASDRDGSNAVYHYSIMSNGAR	444
Db	530	VHIVLVEEDNNAPQFSSEKRYVVOQREBVTGAPLRYATASDRDGSNAVYHYSIMSNGAR	589
QY	445	GOFLDQACGLADVNSPLDYETTKETYLRYADQDGRPPLSNVSGLVTVOYLIDINDAPI	504
Db	590	GHEFIDIBLTGBIQQVAPLADPEAREEYMLARADQAGRPPLSNNTGLASIQVWDINDHPI	649
QY	505	FVSPFPATVLESPLGLYVLVHVOADADAGDNABLEYRLAGVGHDPFETINNGTMSV	564
Db	650	FVSPFPATVLESPLGLYVLVHVOADADAGDNABLEYRLAGVGHDPFETINNGTMSV	709
QY	565	AAELDREBVDYFSGVEARDHGTGTPALTASASVTVLVDVNDNNPTPOPEYTVLANDEAA	624
Db	710	SGPLDRSEVHEHYFVGVEARDHGSPPLSASASVTVTVLVDVNDNRPBFTMKETHYLRINDEAA	769
QY	625	VGTSVTVVSAVDADAHSVITYQTTSQNTRRFESITSGSGGGLVSLALPLDYKLEQRYVLA	684
Db	770	VGTSVTVVSAVDADAHSVITYQTTSQNTRRFESITSGSGGGLVSLALPLDYKLEQRYVLA	829
QY	685	VTASDGTROQTACIQVNVVTDANTHRPVPQSHYTVNVNDRPAGTVTVLSATDEBTGEN	744
Db	830	LITASDRALHCHCVHYNITDANTHRPFGSAHYSVSNEDRPMGSTIIVLSASDDVGEN	889
QY	745	ARITFEMEDSIPOFRIDADTGAVTQAELDYEDQVSYTLATARDNGIPKSDTYLEIL	804
Db	890	ARITYLEEDMIPQFRIDADSGATLQAPLDYEDQVTVTLATARDNGIPKADTYVEVM	949
QY	805	VNVNDNAPQFLADSVQSYEDVPPPTSTLOISANDRDSGLNGRPFYTTQGGDDDDGDP	864
Db	950	VNVNDNAPQFLADSVQSYEDVPPPTSTLOISANDRDSGLNGRPFYTTQGGDDDDGDP	1009
QY	865	IVESTSGIVTTLRLRDRENAVQVYLBAVAADKMPAKRTPEMTVTVLVDVNDMPVPEOD	924
Db	1010	TIBTSGIVTVARRLDRBNAVSVELRYANDREVPLRLTFVSQVWQVNDNAPAPPPAE	1069
QY	925	EPDYFVHEBSPITGLANARVTAIDPDBGTNAQIMYQIVEGNIPEVFOLDIFSGELTALVDL	984
Db	1070	EFEFRVKEBNSIVQSVVAQITAVDPDGGPNAHIMYQIVEGNIPELFQMDIFSGELTALIDL	1129
QY	985	DYEDRPEYVTVIQAQTSAPLSRATVYVRLDNRDNDPVLGNFELFNNTYTNSSSPGG	1044
Db	1130	DYARQRYVTVVQAQTSAPLSRATVYVRLDNRDNDPVLGNFELFNNTYTNSSSPGG	1189
QY	1045	AIGRVPAHDIDISLTYFSEBGNBLSTVLLANASTGELKLSRALDNNRPLEAIVSLVSD	1104
Db	1190	IIGRIPAYDPDVBDELTYFSEBGNBLSTVLLANASTGELKLSRALDNNRPLEAIVSLVSD	1249
QY	1105	GHSVTVQAQCLRTTITTDENLTHSITLRLKEDNSPERFLSPGLFLIOAVAAITLATPPDHV	1164
Db	1250	GHSVTVQAQCLRTTITTDENLTHSITLRLKEDNSPERFLSPGLFLIOAVAAITLATPPDHV	1309
QY	1165	VNVNORDTDAPOGHITLANSLSVGGPPGPGGP--PFLSEBDQEBLYLNRSLTLTISAQ	1222
Db	1310	FITINQIDTDV--GGTYLANSFSLAARAGAGAGAGWPFSSBEIQBOUTYRRALALARSLL	1368
QY	1223	RVLPEPDNICLAREPCENYMCVSVLRFDSAPPIASSVYLRPIHVPVGLRCRCPGFTG	1282
Db	1369	DVLPFPDNDVCLAREPCENYMCVSVLRFDSAPPIASSVYLRPIHVPVGLRCRCPGFTG	1428
QY	1283	DVGETVVDLCYASRPCOPHRCRREBGGYCLCDGYTGHGICVSAASGCTPVCNNGCT	1342
Db	1429	DVGETVVDLCYASRPCOPHRCRREBGGYCLCDGYTGHGICVSAASGCTPVCNNGCT	1488
QY	1343	CNVLLVGFCKDCDPSGD--PEKPYCOVTTTSPFAHSEITFRGLRQREHPTLALSFAKRE	1401
Db	1489	CTPAPNGCFRCQCPAGAGFEGPRCBAVAAFPSSVFMFRGLRQREHPTLALSFAVTVQS	1548
QY	1402	GLLLNNGRPNRKKDPAALSVIQEQVOLTSSAGSTTVVSPFVPGVSDQGMHVOUKYTN	1461
Db	1549	GLLFYNGRINRKKDPAALSVIQEQVOLTSSAGSTTVVSPFVPGVSDQGMHVOUKYTN	1608

QY 1462 KPLIGTGTLPGSEBQKAVAVTVDGCDTGVALRFGSVLGNYSCAAQGTGGSKSLDLTG 1521  
 DB 1609 KPRFDALGAGQSPKDXAVLSTVDDCVAVALPGABIGNYSCAAQVQSSKSLDLTG 1668  
 QY 1522 PLLIGVBDLPBSPFVPMROFVGCRLQVDSRIIDMADPLANNGTYPGCPAKKNVCDN 1581  
 DB 1669 PLLIGVBNLPENFVPHKDPFGCMRDLHI DGRVDMAAFPANNGTAAQCAKHLFDCSG 1728  
 QY 1582 TCHNGTICVNMWDAPCECPJLFGSKSCAQMANNPOHFLGSLVAMH-GLSLPIQPMYL 1640  
 DB 1729 PCNRSGCSERMSFSCDCPFGGKDCQLTMALPHFRNGITLSMFPGDMAVSVMYL 1788  
 QY 1641 SLMFRTQADGVLLQALITRGSTITTLQJLRBGVNL-SV-EGTGLQASSLRLEPGRANDGW 1699  
 DB 1789 GLAFRTATQGVLMQVQAGPHSTLLCQJLRGLLSVTYTRGSG-RASHLLDQVTVSDGRV 1847  
 QY 1700 HMAQLAL--GASGPRGAIL--SPDYQQAABEGMLGRHLGHLNSNTYVGT-PGPRAGV 1754  
 DB 1848 HDLRLLEQEBGRRGHVLMVSLDFSLPQDTMAVSELOGLKVKQLHVGLPGSABEA 1907  
 QY 1755 ARGFRGCTQGVAVSDTEGVNSLDPSHGESINVEQGSLLPDCSNPCPANSYCSNDWS 1814  
 DB 1908 PGLVGCIG-----PPS-----ECGPGCVTNACASGFCPRHADCRLMQT 1948  
 QY 1815 YSCGCDPXYIGDNCTVNCILNPSCHQSVCTRKPSAPHYTCCECPNYLGFYCETRIDQC 1874  
 DB 1949 FSCCTCGGYGPGCVADCLNFCQNGSCRLPQAPHYTCDCGVGFHHCERHMQC 2008  
 QY 1875 PRGMWGHPTGPGCNCDSKGPDPCKNTSGECHEKHNHVRPQSPFTLLDCDCTPTGSLR 1934  
 DB 2009 PRGMWSPTGPGCNCDSVHKGPDPCKNTGQCKEHEHYRPRGSDCLPDCDCTPVGSTR 2068  
 QY 1935 VCDPEBQCEPKGVIGRQCDRCNPPAEVYTNCEVYNVSCPRALBAGIWMPTREGLP 1994  
 DB 2069 SCAPHSGQCCRGALGRQCNCSOSPFAVYTAASCRLYDACRKSLSGVMWQTKRGVL 2128  
 QY 1995 AAAPCPKGSF-----GTAVHCDHEKGMPLPMLFNCTSTPFSBLKFAERLQNNESGLDS 2049  
 DB 2129 ATVPCEPRGALGRGAGAVRLCDEAOGMLBPDLENCTSPAFRELSLLDGLKLNKTLADT 2188  
 QY 2050 GRSQQLALLRNATOFHAGYFGSDPVKAYAYOLATRLAHNESTORGFLSATODVHTENIL 2109  
 DB 2189 MEAKKLAQRIBETGHTDHYFSQDVRYTARLLAHLAFESHQCGFLTATQDAHFNENIL 2248  
 QY 2110 RVGSALLDTRANKHWEIL-QOTBG--GTAMLQHYEAVASALAKNNHTYLSPTIYTP 2165  
 DB 2249 WAGSALLAPETGDMALGQAPGSPGSAGLVRHLEBYAATLARNNELTYLPMGLVTP 2308  
 QY 2166 NIVISVRLDKNF--AGAKLPYEA--LRGEBPDLLETTYILPESVRETPRVVRPAGP 2221  
 DB 2309 NIMLSIDRMHPPSPRGARPRYHNSLFRGODAMPHTVLLPSQSPRPSRSLVPTS 2368  
 QY 2222 GBAQEPBELARQRHRELQGBAVASVITYRTLGLLPHNYDPDKSLVPRKPIINTP 2281  
 DB 2369 SIENSTSSVPPRPAPEPPRPGISITILLVYRILGILLPQFOERKGRARLPQNPVWNSP 2428  
 QY 2282 VSIISVHDEBELLPRALDKPVTVQFLLLEETERTKPICVFNHNSILVSGTGWMSARGCEV 2341  
 DB 2429 VVSVAVHGHANPLRGLIESPISLEFRLQYANRSKALCYQMDPPGLAEQHWATARDCEL 2488  
 QY 2342 VPRNESHVSCQNMHSFAYLMDVSRRE--NGBILPKYLTLYVALGYTLAALLTPEFLT 2399  
 DB 2489 VHRNGSHARCRCSTGTGFMADMASPRERLEGLLELLAVFTHVAVASVALVTLTAAILL 2548  
 QY 2400 LNRILRSNQHIGIRNLTALALQVLTLGINQADLPFACTVAILLHFLYLCFESWALL 2459  
 DB 2549 SLASIKSNVAGIHANVAALGVABELFLGLHGHTHNQVCTAVAILLHYFFLSTPAMLY 2608  
 QY 2460 BALHLYALTEVDVNTGPMKFTYMLGMVPAFTTGLAVGLDBEGYGNPDCWLSIYDTL 2519  
 DB 2609 QGHLTYMOVPRVNDGAMRFYHALGMVPAVLLGLAVGLDBEGYGNPDCWISVHEPL 2668  
 QY 2520 IWSRPAVAVAVSNVFLYTLAABASCA-AQROGFEKKGPVSGLOPSFAVILLLSATWLL 2578

DB 2669 IWSRPAVAVLVIMNGTMTFLAARTSCSTGQREA--KTSALTLRSFILLLVSSWLP 2726  
 QY 2579 ALLSVNSDTLLFHYLFATCNCIQGPFTFLSYVLLSKEVRKALKAC-SKSPSPDPLTTK 2637  
 DB 2727 GLAVHNSILAFHYLAFAGCGIGLAVLLLPCLNADAAAMWPACTIGRKAAPREBARPAV 2786  
 QY 2638 STLSYNGCPSPYADRLYO-PYGDAGSLHSTRSGKSQP-----SYI--PFLREE 2687  
 DB 2787 GLGPAVNTALPESBGLRITLGASTVSSVASBSGRTODDSDQGRSITLRONTLVRG 2846  
 QY 2688 SA-----LMPQGPBGIG-----DPGSLFLEGGDQHDPTDSDSLSEDDGSGY 2734  
 DB 2847 SAADHTDHLQAHAGPTDLVAMFHRDAGA-----DSDESDLSLEBERSLSI 2894  
 QY 2735 ASTHSDSBEBEBEBEBEBEAAFGEOGMBDSLPGARLPLASTPRQGGGPGKAPPG-- 2792  
 DB 2895 PSSESDNGRTGRVQRPICRAAQ-----SERLLTH--PKVDGNDLLSYVPALG 2942  
 QY 2793 -----DPGT-----TAKESSGNGAPEERLENGDALSRGSLGPLPGSSAOPH 2835  
 DB 2943 ECAAPCALQWTGSEBRGLDTSKQANNOQDPAL-TSGD-----FTSL-----GRAQOR 2993  
 QY 2836 KGLKKKCLPTISEKSSLRLP-----LEOCTGSSRGSSA-----SEGSRG 2876  
 DB 2994 KGLKKR-----LQYPLVPTQTRGAPELSWCBAAATLGHRAVPAASVGRIVAGGT 3042  
 QY 2877 GPPRPPRQSLQEOULN 2893  
 DB 3043 GLSLOPASRYSRROLD 3059

Search completed: February 11, 2004, 15:54:56  
 Job time : 74 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 11, 2004, 15:53:34 / Search time 31 Seconds  
(without alignments)  
3989.504 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545

Sequence: 1 MRSPTVGVLPPTPPPLLL.....AGTVDSDSGSEFLFFNLH 2923

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents, AA:  
1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1004.5	6.5	884	2	US-08-465-976A-2
2	1004.5	6.5	884	2	US-08-982-412-2
3	846	5.4	1026	1	US-07-998-003A-95
4	846	5.4	1026	1	US-08-453-274B-95
5	846	5.4	1026	1	US-08-453-695A-95
6	846	5.4	1026	1	US-08-453-161A-95
7	846	5.4	1026	2	US-08-453-702A-95
8	846	5.4	1026	3	US-09-099-639-95
9	846	5.4	1026	5	PCT-US93-12588-95
10	846	5.4	1026	5	PCT-US95-08071-95
11	846	5.4	1203	1	US-07-998-003A-103
12	846	5.4	1203	1	US-08-453-274B-103
13	846	5.4	1203	1	US-08-453-695A-103
14	846	5.4	1203	1	US-08-268-161A-103
15	846	5.4	1203	2	US-08-453-702A-103
16	846	5.4	1203	3	US-09-099-639-103
17	846	5.4	1203	5	PCT-US93-12588-103
18	846	5.4	1203	5	PCT-US95-08071-103
19	735.5	4.7	1466	4	US-09-262-537-58
20	727	4.7	1466	4	US-09-262-537-20
21	719	4.6	1403	4	US-09-262-537-6
22	711.5	4.6	780	4	US-09-262-537-10
23	711.5	4.6	1114	4	US-09-262-537-34
24	711.5	4.6	1177	4	US-09-262-537-2
25	691	4.4	1123	4	US-09-262-537-4
26	628.5	4.0	797	1	US-08-453-695A-112
27	628.5	4.0	797	1	US-08-268-161A-112

28	628.5	4.0	797	2	US-08-453-702A-112	Sequence 112, App
29	628.5	4.0	797	3	PCT-US95-08071-112	Sequence 112, App
30	628.5	4.0	797	5	PCT-US95-08071-112	Sequence 112, App
31	624.5	4.0	787	1	US-08-453-695A-110	Sequence 110, App
32	624.5	4.0	787	1	US-08-268-161A-110	Sequence 110, App
33	624.5	4.0	787	2	US-08-453-702A-110	Sequence 110, App
34	624.5	4.0	787	3	US-09-099-639-110	Sequence 110, App
35	624.5	4.0	787	5	PCT-US95-08071-110	Sequence 110, App
36	611.5	3.9	836	1	US-07-998-003A-105	Sequence 105, App
37	611.5	3.9	836	1	US-08-453-274B-105	Sequence 105, App
38	611.5	3.9	836	1	US-08-453-695A-105	Sequence 105, App
39	611.5	3.9	836	1	US-08-268-161A-105	Sequence 105, App
40	611.5	3.9	836	2	US-08-453-702A-105	Sequence 105, App
41	611.5	3.9	836	3	US-09-099-639-105	Sequence 105, App
42	611.5	3.9	836	5	PCT-US93-12588-105	Sequence 105, App
43	611.5	3.9	836	5	PCT-US95-08071-105	Sequence 105, App
44	611.5	3.9	904	1	US-07-998-003A-97	Sequence 97, App1
45	611.5	3.9	904	1	US-08-453-274B-97	Sequence 97, App1

## ALIGNMENTS

RESULT 1  
US-08-465-976A-2  
; Sequence 2, Application US/08465976A  
; Patent No. 5869632  
; GENERAL INFORMATION:  
; APPLICANT: SOPPET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: RUBEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CABELLA, BYRNE, BAIN GILFILLAN, CECCHI  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,976A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY F  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-444  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 994-1700  
; TELEFAX: (201) 994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 884 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-465-976A-2

Query Match 6.5%; Score 1004.5; DB 2; Length 884;  
Best Local Similarity 37.3%; Pred. No. 3e-63;  
Matches 254; Conservative 103; Mismatches 213; Indels 111; Gaps 20;  
CY 2278 INTVSVISVHDBELLPRALDKVTVQFRLFRERTKPICVPMNSILVSGGGSAR 2337  
DB 1 MRSPTVGVLPPTPPPLLL.....AGTVDSDSGSEFLFFNLH 2923



```

QY 2338 GCEVFRNESHVSCQCNHMTSPAVLMDVSRRE--NGEILPLKTLTYVALGVTALALLTF 2395
DB 61 DCELVHRNGSHARCRCSTRTGTFGLMDASPRERLEBDLELAVFTHVVAVALVLT 120
QY 2396 FFLTLRLRSNQHGRRLNLTALGLAQLVFLGINQADLPACTVIALHLHYLCTFS 2455
DB 121 AILLSRLSKSVNRGIHANVAALGVABELFLGIHRTNQVCTAVAILHLHYLCTFS 180
QY 2456 WALLEBALHYRLTEVRDVTNGPMPRTYMLGMGVPAFTGLAVGLDPGYGNDPCWLSI 2515
DB 181 WLFVQGLHYRQVPEPRVDNGMRFTHALGMGVPAFTGLAVGLDPGYGNDPCWLSI 240
QY 2516 YDTLWSPAGPVAFVAVNSVFLYIILABASCA--AQROGFEKKGPVSGLOPSPAVILLLSA 2574
DB 241 HEPLWSPAGPVAVLYVWNGMTFLIARISCTGOREA--KTSALTLLRSSFLILLVSA 298
QY 2575 TWLALLSVNSDTLFLHYLPATCNCIQGPFTFLSYVLSKEYRKALKLAC--SRKSPDPA 2633
DB 299 SWLFGILLAVNHSILAFHYLHAGLCGLAVLLFCVLNADAAAMMPACIGRKAPEBA 358
QY 2634 LTTKSTLTSYNCPSPRYADGRLYQ--PYGDSAGSLHSTRSGSKOP-----SYL--PFL 2683
DB 359 RPAPGLGPGAYNNTALFEBSGLIRITLGASTVSVSSASGRTOQDSQGRSYLRDNYL 418
QY 2684 LREESA-----LNPQGGPPGLG-----DPGSLFLEGDDQDHPDPTDSDSLSEDDQ 2730
DB 419 VHRGSAADHTDHSLOAHAGPTDLDVDMFHRDAGA-----DSDSDSLSEEBR 466
QY 2731 SGSYASTHSDSEEBEEREEBAAPGEGMDSLGPAERLPLHSTPKDGGPGPKAPW 2790
DB 467 SLSTPSESEDNRTGRFQPLCRAGQ-----SERLTLT--PKVDGNDLLSTW 514
QY 2791 PG-----DGT-----TAKESGNGAPBERLRBNGDALSRGSLGRLPGSS 2831
DB 515 PALGCEAAPCALQWGSERRLGLDTSKDAANNQPDAL--TSGD-----ETSL-----GRA 565
QY 2832 ACPHKKILKKCLPTISEKSLRLP-----LEOCTGSSRSGSA-----SE 2872
DB 566 QQRKGIILKNR-----LQYPLVPQTRGAPELSMCRATLGHRAVPAASYGRITYA 614
QY 2873 GSRGPPPPPPROSLQROLN 2893
DB 615 GCGTSLSQPASRYSSREOLD 635

RESULT 2
US-08-982-412-2
; Sequence 2, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE,
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982.412
; FILING DATE:
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PFI81PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-982-412-2

Query Match 6.5%; Score 1004.5; DB 2; Length 884;
Best Local Similarity 37.3%; Pred. No. 3e-63;
Matches 254; Conservative 103; Mismatches 213; Indels 111; Gaps 20;

QY 2278 INTPVVISYHDEBELPRALDKPVTVQPRLLTEERTKPICVFNHSLIVSGTGWGAR 2337
DB 1 MNSPVSAVAFPHGNFLLGILSPISLEFRLLQTNRSKALCVQMDPGLBQHGVMYAR 60
QY 2338 GCEVFRNESHVSCQCNHMTSPAVLMDVSRRE--NGEILPLKTLTYVALGVTALALLTF 2395
DB 61 DCELVHRNGSHARCRCSTRTGTFGLMDASPRERLEBDLELAVFTHVVAVALVLT 120
QY 2396 FFLTLRLRSNQHGRRLNLTALGLAQLVFLGINQADLPACTVIALHLHYLCTFS 2455
DB 121 AILLSRLSKSVNRGIHANVAALGVABELFLGIHRTNQVCTAVAILHLHYLCTFS 180
QY 2456 WALLEBALHYRLTEVRDVTNGPMPRTYMLGMGVPAFTGLAVGLDPGYGNDPCWLSI 2515
DB 181 WLFVQGLHYRQVPEPRVDNGMRFTHALGMGVPAFTGLAVGLDPGYGNDPCWLSI 240
QY 2516 YDTLWSPAGPVAFVAVNSVFLYIILABASCA--AQROGFEKKGPVSGLOPSPAVILLLSA 2574
DB 241 HEPLWSPAGPVAVLYVWNGMTFLIARISCTGOREA--KTSALTLLRSSFLILLVSA 298
QY 2575 TWLALLSVNSDTLFLHYLPATCNCIQGPFTFLSYVLSKEYRKALKLAC--SRKSPDPA 2633
DB 299 SWLFGILLAVNHSILAFHYLHAGLCGLAVLLFCVLNADAAAMMPACIGRKAPEBA 358
QY 2634 LTTKSTLTSYNCPSPRYADGRLYQ--PYGDSAGSLHSTRSGSKOP-----SYL--PFL 2683
DB 359 RPAPGLGPGAYNNTALFEBSGLIRITLGASTVSVSSASGRTOQDSQGRSYLRDNYL 418
QY 2684 LREESA-----LNPQGGPPGLG-----DPGSLFLEGDDQDHPDPTDSDSLSEDDQ 2730
DB 419 VHRGSAADHTDHSLOAHAGPTDLDVDMFHRDAGA-----DSDSDSLSEEBR 466
QY 2731 SGSYASTHSDSEEBEEREEBAAPGEGMDSLGPAERLPLHSTPKDGGPGPKAPW 2790
DB 467 SLSTPSESEDNRTGRFQPLCRAGQ-----SERLTLT--PKVDGNDLLSTW 514
QY 2791 PG-----DGT-----TAKESGNGAPBERLRBNGDALSRGSLGRLPGSS 2831
DB 515 PALGCEAAPCALQWGSERRLGLDTSKDAANNQPDAL--TSGD-----ETSL-----GRA 565
QY 2832 ACPHKKILKKCLPTISEKSLRLP-----LEOCTGSSRSGSA-----SE 2872
DB 566 QQRKGIILKNR-----LQYPLVPQTRGAPELSMCRATLGHRAVPAASYGRITYA 614
QY 2873 GSRGPPPPPPROSLQROLN 2893
DB 615 GCGTSLSQPASRYSSREOLD 635

RESULT 3
US-07-998-003A-95
; Sequence 95, Application US/07998003A
; Patent No. 5643781
; GENERAL INFORMATION:

```

APPLICANT: Suzuki, Shintaro  
 TITLE OF INVENTION: Protocadherin Materials and Methods  
 NUMBER OF SEQUENCES: 107  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ADDRESSER: Bicknell  
 STREET: 20 South Clark Street  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/998,003A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 564378land, Greta E.  
 REGISTRATION NUMBER: 35,302  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/346-5750  
 TELEFAX: 312/984-9740  
 TELETYPE: 25-3856  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-998-003A-95

Query Match 5.4%; Score 846; DB 1; Length 1026;  
 Best Local Similarity 29.1%; Pred. No. 1e-51;  
 Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVATGDADPPNANI-----LYRLBSGGSPSEVFEIDPRSG-VI 349  
 DB 33 GHATRVVYKVBEPQPTLIGSLADYGPDPVGHLYKLEVG-----APYLAVDKGTGIF 87  
 QY 350 RTRGVDREEVSYQ-----LTVASDQGRDPGRSTTAAVFLSVEDDNDNAPO 398  
 DB 88 TTEISIDEGLEBCQNLFGDPCLIEFEVSTIDLVQNASPRLEQOI--EVQDINDNTFN 145  
 QY 399 FSEKRYVQVREDVTPGAFVLRVTASDRDKGSNAVVHYSIMSGNARGQFYLDAGTALDV 458  
 DB 146 PASPVITLAIPEMTWIGSLFPIPLASDRDAGNGVAYELQVAEDQ-----BEKPOLIV 200  
 QY 459 VSPDIYETTKETTLKVRADQGRPPISNVSLVTVQVLDINDNAPIFVSTPPQATVLSV 518  
 DB 201 MGNLDRERWDSDLTIKQDQGSPPRA--TSALLRVTVLDTNNAPKFERPSEYAELESENS 259  
 QY 519 PLGYLVLAHQALDADAGNARLEVRAGVGHDP-----FTINNGTGMISAAELDERE 572  
 DB 260 PIGHSVIOVANDSDQGANARIEITF---HQAPEVRRLLRLDRNTGLITVQGVNDESD 315  
 QY 573 VDFYSGVEARDHGTPALTFASASVSTVLVDVNDNPTF-----TOPEYTVRLNEDAA 624  
 DB 316 LSTLAFSVLAKDRGNPKSARAAQVVTVVDNDNMTPTIIRIGLVTHQDGMANISEDA 375  
 QY 625 VGTSVVTVSAVDRD--AHSVITYQITSGTNRPRFSITSGSGG-----LVSLALPLDYKL 677  
 DB 316 ERTAAALVQVSDRDEBENAAVTC--VVAQVPPQLRQASSTGSDSKKXFLQTTTPIIDYEX 434  
 QY 678 BRQYVLAVAASDGR---ODTAQIVVNTDANTHRVPPQSXYTVNVNEDRAGATTVLI 734  
 DB 435 VQDYITIEIYAVDSGNPPLASTNSLKVQVVDNDNAPVFQSTGEVAFPEANKEGVYIAEI 494  
 QY 735 SATDEDTGENARITYFMEDSIPO-----FRIDADGAVTTOAEILDYEDQVSYTLAITAR 788

DB 495 TASHADSSNMLVYSLE---PEPAKGLFTISPETGSIQVKTSLDRQRSEYELKAVAA 551  
 QY 789 DNGIPQSDTTYLILVNDVNDNAPQFLRDSYQGSYEDVPEFTSVLOISATDRDSGNG 848  
 DB 552 DRGFSILOGTATVAVNLDCNDNDPKFMLSQYNSVNMENMPLSPVGAVTVIIDDGKENA 611  
 QY 849 RVFTYFGQDGDGDDPFVYESGIVRLRLDRNVNAQVLRAYAVDGMPPARTPMVET 908  
 DB 612 QVQLSV---EODNQPVTYQNGTGTILSLSPDRQOQSTYTQLKAVDGVPPRSAYVET 668  
 QY 909 VTVLDVNDNPPVF--BODEPDEVENSPIGIAVARVATDPDQGTNAQIMQIVEGNIP 966  
 DB 669 INVLDENDNAPIYIAPNSTSHKLLTPQTRIGETVSQVAAEDPDSGVNAELIYSLAGNIPY 728  
 QY 967 EVFQDLDFSGELTALVDLDYDRPEYVLYVIOATAPLVSRAVTRRLDRNDNPPVLN- 1025  
 DB 729 GLPQIGSHSGAITTEKEIERRHGHRLV-----VKVSDRG-KPPRYGTA 772  
 QY 1026 -FEILFNMYVNRSSPPFGAIGRPAH-----DPDISDLTY--SPERENEL 1070  
 DB 773 LVHLVNETLANRK-----LLETLLGHSUDTPIDIDLAGDPYERSKQGNIL 820

RESULT 4  
 US-08-453-274B-95  
 Sequence 95, Application US/08453274B  
 Patent No. 3663500

GENERAL INFORMATION:  
 APPLICANT: Suzuki, Shintaro  
 TITLE OF INVENTION: Protocadherin Materials and Methods  
 NUMBER OF SEQUENCES: 107  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/453,274B  
 FILING DATE: 30-MAY-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5663300and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32660  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELETYPE: 25-3856  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-453-274B-95

Query Match 5.4%; Score 846; DB 1; Length 1026;  
 Best Local Similarity 29.1%; Pred. No. 1e-51;  
 Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVATGDADPPNANI-----LYRLBSGGSPSEVFEIDPRSG-VI 349  
 DB 33 GHATRVVYKVBEPQPTLIGSLADYGPDPVGHLYKLEVG-----APYLAVDKGTGIF 87  
 QY 350 RTRGVDREEVSYQ-----LTVASDQGRDPGRSTTAAVFLSVEDDNDNAPO 398

```

Db      88 TTETSIDREGLECONQOLPDPDCLIEFEVISTDLVONASPRLEBQI--EVODINDINTEN 145
Qy      399 FSEKRYVVOVREDVTPGAPVLRTVATSDRDGKSNAAVHYISIMSGNARGOFLDQOTGALDY 458
Db      146 FASPVITTLAIPENTNIGSLFPIPLASDRDAGPVGASVELQVAEDQ-----BEKOPOLIV 200
Qy      459 VSPIDVETTKETTLRVAQDGGRRPPLSNVGLVTVQVLDINDNAPLFTVSTPQATVLESV 518
Db      201 MGNLDRERWDSYDLTIKVQDGGSPRA--TSALLRVTLVDINDNAPKERSYEAEISENS 259
Qy      519 PLGVTLVHQAIDADAGDARLEFRLAGVHDPF-----FTINNGTGMISVAALDREE 572
Db      260 PIGHSVIOVKANDSOGANAEIEYTF---HOAPEVRRLLRLDNTGILTVQGPVDRD 315
Qy      573 VDFYSGEVRARDGTPLRTASASVTVLVDNNDNPTF-----TOPEYTVRLNEDAA 624
Db      316 LSTLRPSVLAKDGTNPKSARAQVTVKMDNAPLIEIRIGLVTHQDGMANISEDA 375
Qy      625 VGTSVTVASAVDRD--AHSVITYQITSGNTRNRPISITSGGG---LVSLALPLDYKL 677
Db      376 EETAVALVQVSDRDEGENAAVTC--VVAGDVPOLRQASETSGDSKKKYFLQTTPLDYEK 434
Qy      678 ERQYVLAVTASDGR---QDTAQIVNVNTDANTHRPVQSSHYTVANNEDRPACTTVLI 734
Db      435 VKDYTLIEIVAVDSGNPPLSTNSLKVQVVDNNDAPVFTQSVTEVAPENNKGEVIAEI 494
Qy      735 SATDEDTGENARITYFMEDSIPQ-----FRIDADTGAVTTOABLDYEDQVSYTLAITAR 788
Db      495 TADSDAGSGNAELVYSLE---PEPAKGLFTISPGEIOVKTSLDRQESIEELKVAA 551
Qy      789 DNGIPOKSDTTLIELVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLOISATDRDGLNG 848
Db      552 DRGSPLOGTATVAVLWLDNDNDPKFMLSQNFVSMENMPALSPVGMVTVIIGDKGENA 611
Qy      849 RVFYTFQGGDDGDDPIVESTSGIVRTLRLDRENVAQVYLRAVADKGMPPARTPMEYT 908
Db      612 QVQLSV---EQDNGDFVIONGTITLSISLSPRBOQSTYTFOLKAVDGVPPRSAYVGT 668
Qy      909 VTVLDVNDNPPVF--EODEDFVFEENSPIGLAVARTATDPDEGTAQIMYOIVBGNIP 966
Db      669 INVLDENDNAPYITAPNTSHKLLTPQTRIGETYSQVAEDPDSGNVAILIYSIAGNIP 728
Qy      967 EVFQDLIFSGELTALVDLYEDRPREVLYIQATSAPLVSRATVHRLDRNDNPPVLGN- 1025
Db      729 GLFQIGSHSGAITLEKEIERRHGHLRLV-----VKVSDRG--KPPRYGTA 772
Qy      1026 -FELFNNTVNTSSSPFGAIGRPAN-----DPDISLTY--SFEKGNEL 1070
Db      773 LVHLYNVEITLARI-----LLETGLGSLDTPLDIDIDAGDPYERSKQKGNIL 820

RESULT 5
US-08-453-695A-95
; Sequence 95: Application US/08453695A
; Patent No. 5708143
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; ADDRESS: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/453, 695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-695A-95

Query Match      5.4%; Score 846; DB 1; Length 1026;
Best local Similarity 29.1%; Pred. No. 1e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

Qy      305 GREVLTTRATDGDAPPNANI-----LYRLIEGSGSPSEVFEIDPRSG-VI 349
Db      33 GATRAVTVYKPEEQPNTLIGSLADYGPEDVGHYKLEVG---APYLVDGKTGIDIF 87
Qy      350 RTRGPDREVEVSQY-----LVTEASDQGRDGPRTTAVALFSEVDDNDNAPQ 398
Db      88 TTETSIDREGLECONQOLPDPDCLIEFEVISTDLVONASPRLEBQI--EVODINDINTEN 145
Qy      399 FSEKRYVVOVREDVTPGAPVLRTVATSDRDGKSNAAVHYISIMSGNARGOFLDQOTGALDY 458
Db      146 FASPVITTLAIPENTNIGSLFPIPLASDRDAGPVGASVELQVAEDQ-----BEKOPOLIV 200
Qy      459 VSPIDVETTKETTLRVAQDGGRRPPLSNVGLVTVQVLDINDNAPLFTVSTPQATVLESV 518
Db      201 MGNLDRERWDSYDLTIKVQDGGSPRA--TSALLRVTLVDINDNAPKERSYEAEISENS 259
Qy      519 PLGVTLVHQAIDADAGDARLEFRLAGVHDPF-----FTINNGTGMISVAALDREE 572
Db      260 PIGHSVIOVKANDSOGANAEIEYTF---HOAPEVRRLLRLDNTGILTVQGPVDRD 315
Qy      573 VDFYSGEVRARDGTPLRTASASVTVLVDNNDNPTF-----TOPEYTVRLNEDAA 624
Db      316 LSTLRPSVLAKDGTNPKSARAQVTVKMDNAPLIEIRIGLVTHQDGMANISEDA 375
Qy      625 VGTSVTVASAVDRD--AHSVITYQITSGNTRNRPISITSGGG---LVSLALPLDYKL 677
Db      376 EETAVALVQVSDRDEGENAAVTC--VVAGDVPOLRQASETSGDSKKKYFLQTTPLDYEK 434
Qy      678 ERQYVLAVTASDGR---QDTAQIVNVNTDANTHRPVQSSHYTVANNEDRPACTTVLI 734
Db      435 VKDYTLIEIVAVDSGNPPLSTNSLKVQVVDNNDAPVFTQSVTEVAPENNKGEVIAEI 494
Qy      735 SATDEDTGENARITYFMEDSIPQ-----FRIDADTGAVTTOABLDYEDQVSYTLAITAR 788
Db      495 TADSDAGSGNAELVYSLE---PEPAKGLFTISPGEIOVKTSLDRQESIEELKVAA 551
Qy      789 DNGIPOKSDTTLIELVNDVNDNAPQFLRDSYQGSVYEDVPFTSVLOISATDRDGLNG 848
Db      552 DRGSPLOGTATVAVLWLDNDNDPKFMLSQNFVSMENMPALSPVGMVTVIIGDKGENA 611
Qy      849 RVFYTFQGGDDGDDPIVESTSGIVRTLRLDRENVAQVYLRAVADKGMPPARTPMEYT 908
Db      612 QVQLSV---EQDNGDFVIONGTITLSISLSPRBOQSTYTFOLKAVDGVPPRSAYVGT 668
Qy      909 VTVLDVNDNPPVF--EODEDFVFEENSPIGLAVARTATDPDEGTAQIMYOIVBGNIP 966
Db      669 INVLDENDNAPYITAPNTSHKLLTPQTRIGETYSQVAEDPDSGNVAILIYSIAGNIP 728
Qy      967 EVFQDLIFSGELTALVDLYEDRPREVLYIQATSAPLVSRATVHRLDRNDNPPVLGN- 1025

```

Db 729 GLPQIGSHGATITEKIERRHGHLRLV-----VKYSDRG-KPRRYGTA 772

QY 1026 -FEILFNNVYTNRSSSPFGAIGRVPAH-----DPDISDLTY--SFERGNEI 1070

Db 773 LVHLVYNETLANRT-----LLETLLGSHLDTPLDIDIDAGDEYERSKORGNIL 820

RESULT 6

US-08-268-161A-95

Sequence 95, Application US/08268161A

Patent No. 5798224

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/268,161A

FILING DATE: June 27, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Young J. Suh

REGISTRATION NUMBER: P-41,337

REFERENCE/DOCKET NUMBER: 27866/32149

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 1026 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-268-161A-95

Query Match 5.4%; Score 846; DB 1; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;

Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

QY 305 GYEVLTVRATDGDAPPNANT-----LYRLLEGSGGSPSEVEIDPRSG-VI 349

Db 33 GHATRVYKVPKPEQPPNLTIGSLAAYGPPDVGHLYKLEVG-----APYLRVNGKTKGDI 87

QY 350 RTGPGVDREVESEYQ-----LYVEASDQGRDPGPRSTTAAVFLSVEDNDNNAQ 398

Db 88 TTETSIDREGIARCCQNLPGDPCILFEVYSITDLYVGNASRLLEGQI--EVDQINDNTN 145

QY 399 FSEKRYVOVREHVTEGAPVLRATASDRDKGSNAVVHYSIMSGNARGQTLDAQCALV 458

Db 146 FASPVITTLAIPENTNIGSLPEPLASDRDGPAGVASYTEQVAEDQ-----EKKQPOLIV 200

QY 459 VSPLEJETTKEVTLRVARADGGRPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLESV 518

Db 201 MGNLDERHDSYLTITKVDDGSPFPA--TSALLRVTVLDINDNAPFEPSPYAEISENS 259

QY 519 PLGYLVLAHQALDADAGDNARLEYRLAGVGHDP-----FTINNGTGMISVAELDREE 572

Db 260 PIGHSVIQVANDSDQANAEIETP-----HQAFVVRRLRLDRNTGILITQGVPRDSD 315

QY 573 VPFYSGVABRHGTFALTAASVSVTVLDVNDNNTF-----TQPEYTVRLNEDNA 624

Db 316 LSTLRFSVLAKDRGTNPKSARAQVVTVTKOMNDNAPTIIEIRGILGTVTHQDMANISEDVA 375

QY 625 VGRSVYTVASVDRD--ASHVITYQITSNTRRNPFSTTSGSGG-----LVSLALPLDYKL 677

Db 376 BETVALVGVSDRDEGENNAVTC--VVAQGVPPQLRQASSTGSDSKKCYFQTTTPIDYER 434

QY 678 EROYVLAATVADQTR---ODTAQIVNVATDANTHRPVPQSHYTVNVNEDRPAGTVVLI 724

Db 435 VQDYTIEIYANDSGNPPSLSTSLKQVVDVNDNAVFQSYTEVAFPERNKEGVIAEI 494

QY 735 SATDEDTGENNARTYEMEDSIPO-----FRIDADTGAVTQAELDYEDQVSTYLAITAR 788

Db 495 TASDADSGNAELIVSELE--PEPAKGLFTISPETGEIQVTSILDRQRESEYELKVVA 551

QY 789 DNGIPEKSTTYLEIIVNDVNNAPQFLDSYQSGSYTEVEPPFTSYLQISATDRDGLNG 848

Db 552 DRGSPSLQGTATVTVLVVLDCCNDNDPKFMLSQYNFYSVMENMPALSPVGWTVIDGDKGENA 611

QY 849 RVFYTFQGGDDGDDPIVESTGIVRTIARLDRENVAVYLAAYVDKGMPPARTMEYT 908

Db 612 QVQLSV---EQDNGDPVIONGTITLISLSPREQSTYTFQLKAVDGVPPRSAAVGYT 668

QY 909 VTVLVDNDNPVVF--EQDEFVFEENSPIGLAVAATATDDEGTNAQIMTQIVEGNIP 966

Db 669 INVLDENDNAPYTAPSNSTSHKLLFPQTRLGETVSGVAEDSDGVNAELIYSIAGNIFY 728

QY 967 EYFQDIPSGELTALVDLDYERPEVYVYQATSAVLVSARATVHRLDRNDNPPVLCG- 1025

Db 729 GLPQIGSHGATITEKIERRHGHLRLV-----VKYSDRG-KPRRYGTA 772

QY 1026 -FEILFNNVYTNRSSSPFGAIGRVPAH-----DPDISDLTY--SFERGNEI 1070

Db 773 LVHLVYNETLANRT-----LLETLLGSHLDTPLDIDIDAGDEYERSKORGNIL 820

RESULT 7

US-08-453-702A-95

Sequence 95, Application US/08453702A

Patent No. 5891706

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,702A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5891706and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 32657

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6500

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:

LENGTH: 1026 amino acids

TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-453-702a-95

Query Match 5.4%; Score 846; DB 2; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;  
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

```

QY 305 GYEVLTAVRATDGDAPNNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349
DB 33 GHATRVVYKVPKEQPPNTLIGSLADYGPDPVGHLYKLEVG-----APLRVKGKTDIF 87
QY 350 RTGPGVREVEBSYQ-----LYVEASDQGRDPGPRSTTAAVFLSVEDNDNAPQ 398
DB 88 TTETSIDREGLRQCQNLPGDPCILEPEVISTDLVQNASPRLEBQI--EVQDINDNFTN 145
QY 399 FSEKRYVQVREDVTPGAPVLRATASPRDKGSNAVYHYSIMSGNARQGFILDAQTGALDY 458
DB 146 FASPVITLAIPTNTNIGSLPPIPLASDRDAGPVGVAHYELQVABDQ-----EKKOPOLIV 200
QY 459 VSPIDYETTKETYLRAVRAODGGRPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESV 518
DB 201 MGNLDERMDSYDLTIKVDGSGPPRA--TSALLRVTYLDINDNAKFERPSYAELESENS 259
QY 519 PLGYLVIAVQALDADAGNARLERYLAGVGHDP-----FTINNGTGMISVAABLDREE 572
DB 260 PIGHSVIYQVANDSDQGANAEIETTF---HOAPEVVRRLRLDRNTGLITVQGVDRD 315
QY 573 VDFYSGVGEARHDGTPALTLTASASVTVLDVNDNPTF-----TOPEYTVRLMEDAA 624
DB 316 LSTLRPSVLAKGRTNPKSARAQVAVTVKMDNDNAPTEIRIGIGLVTHODGMANISEDA 375
QY 625 VGTSVTVVASVDRD--AHSVITYQITSQNTNRNFSITSQSGG-----LVSLALPLDYKL 677
DB 376 EETAVVALVOVSDRDEGENAAVTC--VAGADVPPQLQASETGDSKSKKFLQTTPLDYEK 434
QY 678 ERQYVLAVTASDGT---QDTAQIVNVNTDANTHRPVFQSSHYTVANNEDRPAGTTVLI 734
DB 435 VKQYTIETVAVDSGNPLSSTNSLKVQVVDVNDNAPFTQSVTEVAFPENNKKEGVIAEI 494
QY 735 SATDEDTGNGENARTTFMEDSIPO-----FRIDADGAVTTOALEYEDQVSTLAIATR 788
DB 495 TMSDADSGSNAELVYLSL---PEPAKGLFTISPEGEIOVKTSLPREQSESTELKVAA 551
QY 789 DNGIPIKSDTYLEILVNDVNDNAPQFLRDSYQGSYVEDVPFTSYVLOISATDRDGLNG 848
DB 552 DRSPSLQGTATYLVANVLDGNNDPKFMLSINFSVWENMPALSPVGMVTVIDGDGENA 611
QY 849 RVPYTTQGGDGDGDFIVESTSGIVKTLRLDRBNVAQYVLAAYVDKMPARTMEYV 908
DB 612 QVQLSLV---EQDNGDFVITONGTGILLISLSPDREQOSTYTFQLKAVDGGVPRSAVYGV 668
QY 909 VTVLDVNDNPPV--BQDEPDEVVEENSPIGLAVARTADDPREGNAQIMQIVESNIP 966
DB 669 INVLDENDNAPYTTASNTSHKLLTPQTRIGETVSQVAAEDPDSGVAAELIYSIAGENY 728
QY 967 EYFQOLDIFSGELTALVDLYEDREPEVLYIQATSAPLVSRATVHVRLLDRNDPPLYGN- 1025
DB 729 GLPQIGSHSGAILTEKIEIRRHGHLRLV-----VAKSDRG--KDPYRGTA 772
QY 1026 -FELLFNNTYTRSSSPFGAIGRVPAH-----DPLISDLYT--SFERGNEI 1070
DB 773 LVHLVYNETLANRT-----LLETLLGHSIDTPLDIDIDADPEYERSKORGNIL 820

```

## RESULT 8

US-09-639-95  
Sequence 95; Application US/09099639  
Patent No. 6262237  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 115

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/099,639  
FILING DATE: 18 JUN 1998

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,161

FILING DATE: 27 JUN 1994  
ATTORNEY/AGENT INFORMATION:

NAME: Greta B. No. 6262237 and  
REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/34703  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448

TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 95:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1026 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-639-95

Query Match 5.4%; Score 846; DB 3; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51;  
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

```

QY 305 GYEVLTAVRATDGDAPNNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349
DB 33 GHATRVVYKVPKEQPPNTLIGSLADYGPDPVGHLYKLEVG-----APLRVKGKTDIF 87
QY 350 RTGPGVREVEBSYQ-----LYVEASDQGRDPGPRSTTAAVFLSVEDNDNAPQ 398
DB 88 TTETSIDREGLRQCQNLPGDPCILEPEVISTDLVQNASPRLEBQI--EVQDINDNFTN 145
QY 399 FSEKRYVQVREDVTPGAPVLRATASPRDKGSNAVYHYSIMSGNARQGFILDAQTGALDY 458
DB 146 FASPVITLAIPTNTNIGSLPPIPLASDRDAGPVGVAHYELQVABDQ-----EKKOPOLIV 200
QY 459 VSPIDYETTKETYLRAVRAODGGRPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESV 518
DB 201 MGNLDERMDSYDLTIKVDGSGPPRA--TSALLRVTYLDINDNAKFERPSYAELESENS 259
QY 519 PLGYLVIAVQALDADAGNARLERYLAGVGHDP-----FTINNGTGMISVAABLDREE 572
DB 260 PIGHSVIYQVANDSDQGANAEIETTF---HOAPEVVRRLRLDRNTGLITVQGVDRD 315
QY 573 VDFYSGVGEARHDGTPALTLTASASVTVLDVNDNPTF-----TOPEYTVRLMEDAA 624
DB 316 LSTLRPSVLAKGRTNPKSARAQVAVTVKMDNDNAPTEIRIGIGLVTHODGMANISEDA 375
QY 625 VGTSVTVVASVDRD--AHSVITYQITSQNTNRNFSITSQSGG-----LVSLALPLDYKL 677
DB 376 EETAVVALVOVSDRDEGENAAVTC--VAGADVPPQLQASETGDSKSKKFLQTTPLDYEK 434
QY 678 ERQYVLAVTASDGT---QDTAQIVNVNTDANTHRPVFQSSHYTVANNEDRPAGTTVLI 734
DB 435 VKQYTIETVAVDSGNPLSSTNSLKVQVVDVNDNAPFTQSVTEVAFPENNKKEGVIAEI 494
QY 735 SATDEDTGNGENARTTFMEDSIPO-----FRIDADGAVTTOALEYEDQVSTLAIATR 788

```

Db 495 TASDADSGSNALVYSLE---PEPAKGLFTISPERGEIQTSLDREQRSEYELKVAA 551  
 Qy 789 DNGIQRKSDTYLEILVNDVNDNAPQLDSDYQSGVYEDVPFTSVLQISATDRDGLNG 848  
 Db 552 DRGSPSLQGTATVAVLVDLCNDNDPKFMLSGVNFSEWEMPALSPGVMTVIDGKGENA 611  
 Qy 849 RFEYTFQGGDDGDPFVSTSGIVRTLRDLRENAQVLAIVADKMPAPRTPEMYT 908  
 Db 612 QVQLSV---EQDNGDFVIQNGTGTLLSLSFDRBOQSTYTFQLKAVDGGVPPRSAYVGT 668  
 Qy 909 VTLVDNDNPPVFE---EODEFVFEVENSPIGLAVARVATDPEEGNAQIMQIVEGNIP 966  
 Db 669 INVLDENDNAPYITAPNSHTLIPQTRLGETSVQVALEDPSGVNAELIYSIAGNIPY 728  
 Qy 967 EVFQLDIFSGELFALVDLYEDRPEVYVIOATSAPLVSRATVAVRLDNDNPPVGN- 1025  
 Db 729 GLFQIGSHGATILEKEIERRHGHLRLV-----YKVSDRG-KPPRYGTA 772  
 Qy 1026 -FEILFNNYVNNRSSSPFGAIGRPVPAH-----DPDISDILTY--SFERGNEI 1070  
 Db 773 LVHLVYNETLANRT-----LLETLLGHSLDTPDLIDDIADGPEYERSKQGNIL 820

## RESULT 9

PCT-US93-12588-95  
 ; Sequence 95, Application PC/TUS9312588  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 107  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ; ADDRESSER: Borun  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/12588  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/998,003  
 ; FILING DATE: 29 DEC 1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 31811  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 95:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1026 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US93-12588-95

Query Match 5.4%; Score 846; DB 5; Length 1026;

Best Local Similarity 29.1%; Pred. No. 1e-51; Mismatches 346; Indels 114; Gaps 24;

Qy 305 GYEVLTVATGADAPPNANI-----LYRLLESGSGSPSEVFEIIPRSG-VI 349  
 Db 33 GHATRVVVKVPEBQPNLLIGSLADYGFPPDVGHLYKLEVG-----AYLAVADGKTDGIF 87

Qy 350 RTRGVDRREEVSYQ-----LTVASDGRDGPSTTAAPLSEYEDNDNAPQ 398  
 Db 88 TTERSIDREGAREQONQAPGDCILFEVSTLTLVQNASPRLLEBQI--EVQIDINDTFN 145  
 Qy 399 FSEKRYVQVREDYTPGAPVLRTASDRDKGSNAVHYSINSGNARQGFYLDQGTALDV 458  
 Db 146 FASPVITLAIPEENTNIGSLPPIPLASDRDAGPNQVASYELQVAEDQ-----EEKQPOLIV 200  
 Qy 459 VSPLDYETTEKEYTLRVARQDGRGRLPSVSGLVVQVYLIDINDNAPIFSTPFGQATVESV 518  
 Db 201 MANDIREKWDSDYLITKQDGGSPRA-TSALMLVYLTIDNDNAPKERPSYEAELSENS 259  
 Qy 519 PLGYLVHVAQIADADGNALREYLAGVGDPP-----FTINNGTGMISVALEIDREE 572  
 Db 260 PIGHSVIGVANDSDQGNMEIETFP---HQAEVYRRLRLRBNGLIIVQGVDRRED 315  
 Qy 573 VDFYSGVARDHGTPLATASASVTVLVNDNPPF-----TQPEYVRLMEDAA 624  
 Db 316 LSTRFSTLARDGTNPKSARAQVAVVVKMDNDNAPTEIRIGILVTHQDMANISEDA 375  
 Qy 625 VGTSVTVSAVDRD--ASVITYQITSGNTRNRSITSGSGG-----LVSLALPLYKL 677  
 Db 376 ERTAVLVQVSDRDEGENAVTC-VVAGDVPFLRQASETSGSKKTFLOTTPLDYEK 434  
 Qy 678 EROYLVAVTASDGR---ODTAQIVVNTDANTERPVSQSHYTVNVEDRPAQTIVLI 734  
 Db 435 VKDTIETIVANDSNPLPSTNSLKVQVVDNDAPFTQSVTEVAPENNKEGEVIAEI 494  
 Qy 735 SATDEDTGENAKRTYFMEDSIPQ-----FRIDAGVAVTQAEIYEDVDQSYTLAITAR 788  
 Db 495 TASDADSGSNALVYSLE---PEPAKGLFTISPERGEIQTSLDREQRSEYELKVAA 551  
 Qy 789 DNGIQRKSDTYLEILVNDVNDNAPQLDSDYQSGVYEDVPFTSVLQISATDRDGLNG 848  
 Db 552 DRGSPSLQGTATVAVLVDLCNDNDPKFMLSGVNFSEWEMPALSPGVMTVIDGKGENA 611  
 Qy 849 RFEYTFQGGDDGDPFVSTSGIVRTLRDLRENAQVLAIVADKMPAPRTPEMYT 908  
 Db 612 QVQLSV---EQDNGDFVIQNGTGTLLSLSFDRBOQSTYTFQLKAVDGGVPPRSAYVGT 668  
 Qy 909 VTLVDNDNPPVFE---EODEFVFEVENSPIGLAVARVATDPEEGNAQIMQIVEGNIP 966  
 Db 669 INVLDENDNAPYITAPNSHTLIPQTRLGETSVQVALEDPSGVNAELIYSIAGNIPY 728  
 Qy 967 EVFQLDIFSGELFALVDLYEDRPEVYVIOATSAPLVSRATVAVRLDNDNPPVGN- 1025  
 Db 729 GLFQIGSHGATILEKEIERRHGHLRLV-----YKVSDRG-KPPRYGTA 772  
 Qy 1026 -FEILFNNYVNNRSSSPFGAIGRPVPAH-----DPDISDILTY--SFERGNEI 1070  
 Db 773 LVHLVYNETLANRT-----LLETLLGHSLDTPDLIDDIADGPEYERSKQGNIL 820

## RESULT 10

PCT-US95-08071-95  
 ; Sequence 95, Application PC/TUS9508071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Marshall, O'Toole, Gerstein, Murray, &  
 ; ADDRESSER: Borun  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/08071  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/12588  
 FILING DATE: 23 DEC 1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/998,003  
 FILING DATE: 29 DEC 1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Noland, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32149  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELETYPE: 25-3856  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1026 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US95-08071-95

Query Match 5.4%; Score 846; DB 5; Length 1026;  
 Best Local Similarity 29.1%; Pred. No. 1e-51;

Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

305 GYEVLTVRATDGDAPPNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349  
 33 GHATRVVYKVEBQCPNTLIGSLADYGPDPVGHLYKLEVG-----APYLRVDGKTDIF 87  
 350 RTGSPVDREVEVYQ-----LTVEASDQGRDPRGRTTAAVFISVEDDNDNAQ 398  
 88 TTSTSIDREGLRBCNQQLPGDPCILFEFVSITLVONASPRLLGQI--EVQDINNTN 145  
 399 FSEKRYVQVREBVTGAPVLRVTASDRDGSNAVVHYSIMSGNARQGYLDQGTALV 458  
 146 FASPVITLAIPEENTIGSLPPIPLASDRDAGPNGVASYELQVADQ-----EKKQPLIV 200  
 459 VSPLDYETTKETTLKRAVDGGRPPPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLE 518  
 201 MGNLDRERMDSYDLTIKVGQSGSPRA--TSALRATVTLDTNDNAPKFERPSYAELE 259  
 519 PLGIVLTHVQATADAGDNARLEKRLAGVGHDP-----FTINNGTGMISVAEEDRE 572  
 260 PIGHSVIOVANDSDGANAIEYTF---HQPVEVRRLLRLDRNTGLITVQGPVDRD 315  
 573 VDEYSGVEARHDGTPALTFASASVTVLVDNNDNPTF-----TOPEYTVRLNEDAA 624  
 316 LSTLRPSVLAKDGTGPKSRAQVVTVDMDNAPFIERIGLGLVTHQDGMANISEDA 375  
 625 VGTSVTVASVARD--AHSVITYQTSGNTRNFSITSGGG-----LVSIALPLDYKL 677  
 376 EETFAVALVQVSDREBENAVTC--VVAQVDFQLRQASETSGSKKCYFLQTTPLDYRK 434  
 678 ERYVLAVALASDSTR---QDTAIVVAVNDANTHRPPVQSGSHTVVANNEDPRAGTIVLI 734  
 435 VKQYTIETVAVDSDGNPLSTNSLKVQVVDNNAPEFTQSVTEVAPENNNKGEVIARI 494  
 735 SATDEDTGENARITTFWEDSIPO-----FRIDATGAVTTQALVYEDOVSYTLAIFAR 788  
 495 TADADSGSNAELVYSLE---PEPAKGLFTISPERGEIOVKSLSRQEGESTELAVVAA 551  
 789 DNGIPOKSDTTYETLILVNDVNDAPQFLRDSYQSGSYVEDVPPTSVLQISATDRSGLNG 848  
 552 DRGSPSLQGTATVTLVNLDCNDNDPKFMLSNGVFNSENNPALSPIVGMVTVLIDGKGENA 611  
 849 RVTYTCGGDGDGDPFVSTSGIVRTLRLDRENVAAQVYLRVAVADKGMPPARTMENT 908

612 QVALSV---EQDNGDPIVIONGTGTLISLSFDRBQSGTYTFQKAVDGVPPRPAVYGV 668  
 909 VTVLVDNNDNPPVFE--EQEBDFVFEENSPIGLAVARVATATPDBECTNAQIMQVIEGNI 966  
 669 INVLDENNAPYTPASNTSHKLTTPQTRLGRTVSQVAAEPDSDGVNALLIYSIAGNPNY 728  
 967 EVFQDLIFSGELTALVDYEDRPEYVLTQATSAPLVSRAVTHRLDNDNPPVLGN- 1025  
 729 GLFQIGSHSGATLLEKEIERRHGHGRLV-----YKVEDRG-KPRTYGTIA 772  
 1026 -FEILFNNTNRSSSPFGAIGRPVAF-----DPDISLITY--SPERGNEL 1070  
 773 LVHLVYNETLANRT-----LLETILGSLDTPLDIDDIAGBPEYERSKQGNIL 820

RESULT 11  
 US-07-998-003A-103  
 Sequence 103, Application US/07998003A

Patent No. 5643781

GENERAL INFORMATION:  
 APPLICANT: Suzuki, Shuntaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 STREET: 20 South Clark Street  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA

ZIP: 60603

COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/998,003A

FILING DATE:  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5643781and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 30903

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/346-5750

TELEFAX: 312/984-9740

INFORMATION FOR SEQ ID NO: 103:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1203 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

US-07-998-003A-103

Query Match 5.4%; Score 846; DB 1; Length 1203;  
 Best Local Similarity 29.1%; Pred. No. 1.3e-51;

Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

305 GYEVLTVRATDGDAPPNANI-----LYRLLEGSGSGSPSEVEFIDPRSG-VI 349  
 33 GHATRVVYKVEBQCPNTLIGSLADYGPDPVGHLYKLEVG-----APYLRVDGKTDIF 87  
 350 RTGSPVDREVEVYQ-----LTVEASDQGRDPRGRTTAAVFISVEDDNDNAQ 398  
 88 TTSTSIDREGLRBCNQQLPGDPCILFEFVSITLVONASPRLLBQI--EVQDINDNTPN 145  
 399 FSEKRYVQVREBVTGAPVLRVTASDRDGSNAVVHYSIMSGNARQGYLDQGTALDV 458  
 146 FASPVITLAIPEENTIGSLPPIPLASDRDAGPNGVASYELQVADQ-----EKKQPLIV 200  
 459 VSPLDYETTKETTLKRAVDGGRPPPLSNVSGLVTVQVLDINDNAPIFVSTPQATVLE 518







```

Db      33 GHATRVVVKVBEOPNTLIGSLADYGFPPDVGHLYKLEVG-----APYLKVDGKTGDI 87
Qy      350 RTGPGVDEEVEESYQ-----LTVEASDQGRDPCGRSTTAAVFLSVEDNDNAPQ 398
Db      88 TTETSIDREGARECONQJLPCILFEVSTIDLVQNASFLBLEQT--EVQDINDNTN 145
Qy      399 FSEKRYVQVREDVTGAPVLRTASDRDKGSNAVHYISMGNAQGFYLDAGTALDV 458
Db      146 FASPVITLAIPEENTIGSLPPIPLASDRDAGPNGVASYELQVAEDQ-----BEKOPQIIV 200
Qy      459 VSPIDYETTKETYLKVRADGGRPPLSVNGLVTVQVLDINDNAPIFSTPQATVLESV 518
Db      201 MGNLDREKWDSDYLTIKVQDGSSPPRA--TSALLRTVLDINDNAPKFERPSEAELESENS 259
Qy      519 PLGYLVHVAQIADADGNARLEAYLAGVGHDP-----FTINNGTGMISVAELDREE 572
Db      260 PIGHSVIGVKNANDQGNANEIETFF-----HQAPEVVRRLRLDRNTGLITVQGPVRED 315
Qy      573 VDFYSFGVEARDHGTPALRTASASVTVLDVNDNPTF-----TQPEYTVRLMEDAA 624
Db      316 LSTRFSVLAQRGTGTPKSAQAQVVVTVKMDNDNAPTEIRIGLGLVTHQGMANISEDVA 375
Qy      625 VGTSVTVVSAARD--AHSVITQITSGTNRKRSITSGSGG-----LVSLAPLDYKL 677
Db      376 EETVAVLVQVSDRDEGENAAVTC--VVAQDVPPQLRQASSETSDSKKXYLTQTTPLDYEK 434
Qy      678 EROYVLAVTASDGR---QDTAQIVNVNTDANTHRPVQSSHVTVNVEDRPACTTVLI 734
Db      435 VKQDTIEIVAVDSGNPPLSTNSLKVQVVDNDNAPVTQSVTEVAPENKKEGEVIAEI 494
Qy      735 SADEDETGEMARITYMEDSIPQ-----FRIDATGAVTTQABLDEYDQVSYTLAITAR 788
Db      495 TASDADSGSNAELVYSLE---PEPAKGLFTISPETGEIQVKTSLDREGRSYELKVVA 551
Qy      789 DNGIPQSDOTTYLEILVNDVNDNAPQPLADSYQSSVYEDVPPFTSVLQISATDQSLNG 848
Db      552 DRGSPSLQGTATVAVNLDCNDNDPKFMLSINFSVNMENPALSPVGAVTVITDDKEMNA 611
Qy      849 RVFTTPOGDDGDDFIVESTSGIVRTLRLRLRENVAQVLRAYAVDKGMPARTMEVT 908
Db      612 QVQLSV---EQDNQDFVINGTGTILSLSPDEQSGSTYTFQLAKADGQVPRSAVYGT 668
Qy      909 VTVLDVNDNPPVF--EQDEFVVEENSPIGLAVARTATDDEGTNAQIMYQIVBGNIP 966
Db      669 INVLEDDNNAFYITAPSNTSHKLLTPQTRLGRTVSQVAAEDPDSGVNAELIYSIAGNPF 728
Qy      967 EVFPLDIFSGELTALVLDYDRPEYVLVYQATSAPLVSRATVAVRLDRDNDPPVLGN- 1025
Db      729 GLFQIGSHSGAITLEKEIERHHGILHRLV-----VKVSDRG--KEPRYGT 772
Qy      1026 -FELIPNNVYVTRSSSPFGAIGRVAH-----DPDISDLTY--SPERGNEL 1070
Db      773 LVHLYVNETLANR-----LLETILGHSUDPTDIDIDAGPPEYERKQKGNIL 820

```

```

RESULT 15
US-08-453-702A-103
: Sequence 103, Application US/08453702A
: Patent No. 5891706
: GENERAL INFORMATION:
: APPLICANT: Suzuki, Shintaro
: TITLE OF INVENTION: Protocadherin Materials and Methods
: NUMBER OF SEQUENCES: 115
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
: STREET: 233 South Wacker, 6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/453, 702A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5891706and, Greta B.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 32657
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELETYPE: 25-3856
: INFORMATION FOR SEQ ID NO: 103:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1203 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-453-702A-103

Query Match 5.4%; Score 846; DB 2; Length 1203;
Best Local Similarity 29.1%; Pred. No. 1.3e-51;
Matches 243; Conservative 131; Mismatches 346; Indels 114; Gaps 24;

Qy      305 GYEVLYTRADGAPPRANI-----LTRLGSGSGSPBEVEIIPRSG-VI 349
Db      33 GHATRVVVKVBEOPNTLIGSLADYGFPPDVGHLYKLEVG-----APYLKVDGKTGDI 87
Qy      350 RTGPGVDEEVEESYQ-----LTVEASDQGRDPCGRSTTAAVFLSVEDNDNAPQ 398
Db      88 TTETSIDREGARECONQJLPCILFEVSTIDLVQNASFLBLEQT--EVQDINDNTN 145
Qy      399 FSEKRYVQVREDVTGAPVLRTASDRDKGSNAVHYISMGNAQGFYLDAGTALDV 458
Db      146 FASPVITLAIPEENTIGSLPPIPLASDRDAGPNGVASYELQVAEDQ-----BEKOPQIIV 200
Qy      459 VSPIDYETTKETYLKVRADGGRPPLSVNGLVTVQVLDINDNAPIFVSTPQATVLESV 518
Db      201 MGNLDREKWDSDYLTIKVQDGSSPPRA--TSALLRTVLDINDNAPKFERPSEAELESENS 259
Qy      519 PLGYLVHVAQIADADGNARLEAYLAGVGHDP-----FTINNGTGMISVAELDREE 572
Db      260 PIGHSVIGVKNANDQGNANEIETFF-----HQAPEVVRRLRLDRNTGLITVQGPVRED 315
Qy      573 VDFYSFGVEARDHGTPALRTASASVTVLDVNDNPTF-----TQPEYTVRLMEDAA 624
Db      316 LSTRFSVLAQRGTGTPKSAQAQVVVTVKMDNDNAPTEIRIGLGLVTHQGMANISEDVA 375
Qy      625 VGTSVTVVSAVND--AHSVITQITSGTNRKRSITSGSGG-----LVSLAPLDYKL 677
Db      376 EETVAVLVQVSDRDEGENAAVTC--VVAQDVPPQLRQASSETSDSKKXYLTQTTPLDYEK 434
Qy      678 EROYVLAVTASDGR---QDTAQIVNVNTDANTHRPVQSSHVTVNVEDRPACTTVLI 734
Db      435 VKQDTIEIVAVDSGNPPLSTNSLKVQVVDNDNAPVTQSVTEVAPENKKEGEVIAEI 494
Qy      735 SADEDETGEMARITYMEDSIPQ-----FRIDATGAVTTQABLDEYDQVSYTLAITAR 788
Db      495 TASDADSGSNAELVYSLE---PEPAKGLFTISPETGEIQVKTSLDREGRSYELKVVA 551
Qy      789 DNGIPQSDOTTYLEILVNDVNDNAPQPLADSYQSSVYEDVPPFTSVLQISATDQSLNG 848
Db      552 DRGSPSLQGTATVAVNLDCNDNDPKFMLSINFSVNMENPALSPVGAVTVITDDKEMNA 611
Qy      849 RVFTTPOGDDGDDFIVESTSGIVRTLRLRLRENVAQVLRAYAVDKGMPARTMEVT 908
Db      612 QVQLSV---EQDNQDFVINGTGTILSLSPDEQSGSTYTFQLAKADGQVPRSAVYGT 668
Qy      909 VTVLDVNDNPPVF--EQDEFVVEENSPIGLAVARTATDDEGTNAQIMYQIVBGNIP 966

```

```

D6 669 INVDENDMNAVITAPSTSHKLTLPQRLESETYSQVAAEFPDSQVNAELLYSTAGNIPY 728
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 967 EVFOLDIFSGELTALVDLDYDRPEYVIVIOATSAPLVSRAITVHVLLDRNDNPVLGN 1025P
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
D6 729 GLPQIGSHSGAITEKEIERHHGHLRV-----VKVSDRG-KPPRYSTA 772
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1026 -FELLFNPNVYNRSSPFGAIGRVPAH-----DPDISLTY--SPERGNEI 1070
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
D6 773 LVHLYVNETLANRT-----LLETIGLSLTPLDIDIDINAGPPEYERSKQGNIL 820
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Search completed: February 11, 2004, 15:58:26  
Job time : 34 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 17:44:36 ; Search time 22139 Seconds  
(without alignments)  
5401.276 Million cell updates/sec

Title: US-09-916-849A-3  
Perfect score: 15545  
Sequence: 1 MSPPATGVLPPTPPPLLL.....AGTVDEDSGSGSEFLFFNLH 2923

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Pgapop 6.0 , Pgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+pn.model -DBV=x1h  
-O=/cgn2\_1/USPTO/SP0916849/runat\_11022004\_143816\_20938/asp\_query.fasta\_1.3079  
-DB=GenEmbl -QFMT=fastap -SUPRFX=rge -MINMATCH=0 -ICODEXT=0  
-INITs=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCFALIG=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0916849@cgn2\_1 15745 @runat\_11022004\_143816\_20938 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEDUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15545	100.0	8772	6	AX224128 Sequence
2	15545	100.0	10531	6	AX549238 Sequence
3	15545	100.0	10531	9	AF234887 Homo sapi
4	15518.5	99.8	8871	6	AX224126 Sequence
5	15279	98.3	9321	6	AX367102 Sequence
6	14647.5	94.2	8950	10	AB028499 Mus muscu
7	12833	82.6	8924	9	D87469 Human mRNA
8	12339	79.4	24370	6	AX646739 Sequence
9	12339	79.4	24370	9	AB065955 Homo sapi
10	12339	79.4	169241	2	AL390252 Human DNA
11	12339	79.4	227194	2	BX284647 Homo sapi
12	12001	77.2	235320	2	AC106175 Rattus no
13	12001	77.2	235327	2	AC095463 Rattus no
14	11883	77.1	254975	2	AC095125 Rattus no
15	11892	76.5	248059	2	AC113756 Rattus no
16	11783.5	75.8	89129	10	AL672200 Mouse DNA
17	11097	71.4	6699	10	AB011529 Rattus no
18	10517.5	67.7	234932	2	AC093365 Mus muscu
19	9016	58.0	10919	10	AF031572 Mus muscu
20	8991.5	57.8	11660	2	AL157901 Homo sapi
21	8781.5	56.5	9045	6	AX223857 Sequence
22	8781.5	56.5	11389	6	AX549158 Sequence
23	8781.5	56.2	11389	9	AF231024 Homo sapi
24	8736.5	56.2	9087	6	AX226507 Sequence
25	8710	56.0	10195	6	AE0094 Sequence 4
26	7836.5	50.4	11956	9	AF231023 Homo sapi
27	7836.5	50.4	11965	6	AX549370 Sequence
28	7819.5	50.3	12348	6	AX662321 Sequence
29	7780.5	50.1	11868	10	AB011528 Rattus no
30	7766	50.0	11648	6	AX174748 Sequence
31	7736.5	49.8	11541	10	AF427498 Mus muscu
32	5554	35.7	206908	9	AC121252 Homo sapi
33	5457	35.1	88326	9	AC005923 Homo sapi
34	4984.5	32.1	239657	2	AC096455 Rattus no
35	4825	31.0	12634	3	AF172329 Drosophi
36	4783.5	30.8	12200	3	AB028498 Drosophi
37	4766.5	30.7	143393	2	AC015222 Drosophi
38	4766.5	30.7	182494	3	AC009343 Drosophi
39	4766.5	30.7	182494	3	AB003828 Drosophi
40	4651.5	29.9	257909	9	AE0031437 Homo sapi
41	4584.5	29.5	6791	6	AE0092 Sequence 2
42	4453	28.6	2653	10	AF031573 Mus muscu
43	4433	28.5	3343	10	BC005499 Mus muscu
44	3769.5	24.2	3912	6	AX098223 Sequence
45	3602.5	23.2	216773	2	AC116764 Mus muscu

RESULT 1

ALIGNMENTS

AX224128  
LOCUS AX224128 8772 bp DNA linear PAT 10-SEP-2001  
DEFINITION Sequence 3 from Patent WO0161003.  
ACCESSION AX224128  
VERSION AX224128.1 GI:15554400  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 Testa, T.T.  
The human homologue of flamingo  
Patent: WO 0161003-A 3 23-AUG-2001;  
JOURNAL SWITZERLAND BEECHAM PLC (GB)  
FEATURES  
source  
1. 8772  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 1696 a 2814 c 2564 g 1698 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 8772  
Score: 15545.00 Matches: 2923  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-09-916-849a-3 (1-2923) x AX224128 (1-8772)  
QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
Db 1 ATGCGAGGCCGCGCCACCGCGTCCCTCCCAAGCCGCGCGCGCGCTGCTGCTG 60  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
Db 61 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyThrLeuCysProSer 60  
Db 121 GGGTCCAGGGGAGAGAGGCTTTCGGGGGCTGCGCCCAATGAGCTGCTGCTGCTGCT 180  
QY 61 SerAlaSerAsnLeuThrLeuThrGlyThrSerArgCysArgAspAlaGlyThrGlyLeuThr 80  
Db 181 TCAGCGTCCGAACCTTGGCTCTACACAGCCGCTGACAGGATGCGGCACTGAGCTGACT 240  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTTPCysPProGlnSerGlnAlaHis 100  
Db 241 GGGCACTGCTACCCCAACAGATGAGCTGAGGGTTCGATGTCAGAAATCCAGAGCCCAT 300  
QY 101 IleProLeuProProAlaProGlnGlyCysProTTPSerCysArgLeuLeuGlyIleGly 120  
Db 301 ATTCCCTTACACAGAGCTCTGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGlnGlnHisProCysLeuValAla 140  
Db 361 GGGCACTTTCCTCCCAACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
QY 141 ProArgLeuArgCysGlnSerCysLysLeuValGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 421 CCAGCGCTCAGATGCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
QY 161 ArgSerProGlnGlnSerLeuGlyLysArgLysArgAsnValAsnThrAlaProGln 180  
Db 481 AGGTACCAAGAAAGTCCCTGGGTGGGCTGCGAAAGAAAGTAAATACAGCCCCCAG 540  
QY 181 PheGlnProProSerGlyArgAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200  
Db 541 TTCACAGCCCCCAGCTACAGGCGCACAGTCCCGAAGAACAGGACAGGCACTCCCTGTT 600

QY 201 AlaSerLeuArgAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlnThrMet 220  
Db 601 GGAATCCCTTGAAGGCGCATGACCCCGACAGAGGTGAGGAGGTGCAGTGAAGTACACAG 660  
QY 221 AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla 240  
Db 661 GATGCCCTCTTGTATAGCCGCTCCCAACAGTCTTCTCCCTGAGCCAGTCACTGCTGCA 720  
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyThrLysSerThrHisValPheArgValThr 260  
Db 721 GTAAACCAAGCCAGAGAGCTGATCGTAGACCAAGAGCACCGTCTTCAGGGTCAGG 780  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
Db 781 GGGCAGAGACCAAGGATGCCCGAGAGTGCCTTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlyLysArgLysLeuArgGlu 300  
Db 841 GACACCAATGACCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
QY 301 AsnLeuGlnValGlyLysGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 901 AACCTGAGAGTTCGCTATGAGGTGCTCACTGTCAGGGCCAGAGATGATGATCCCTCC 960  
QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340  
Db 961 AATGCCAATATTCGTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
QY 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlnVal 360  
Db 1021 GAGATCGACCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
QY 361 GlnSerGlyGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
Db 1081 GAATCTTACCACTACAGGTAGAGGACAGTACAGAGTCCAGGCTCCGAGTCTTCCGAGT 1140  
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAsnAlaProGlnPheSer 400  
Db 1141 ACCACAGCCGCTGCTTTCCTTCTGCTGAGATGACAAATGATATGATGCTCCCAAGTT 1200  
QY 401 GlnLysArgGlyValValGlnValArgGlnAspValThrProGlnValProValLeuArg 420  
Db 1201 GAGAAAGCTGATGTGTCCAGGTGAGGAGATGATGACTCCAGGGGCCCTCACTACCG 1260  
QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValHisIleIleSerIleMetSer 440  
Db 1261 GTCACAGCTCGAGTCGAGACAGAGGAGCAATGCCGTGTGCTCATATGACTATGACT 1320  
QY 441 GlyAsnAlaArgGlyGlnPheThrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
Db 1321 GGCATATGCTCGGGGAGCAGTTTATCTGATGCCAGACTGGAGCTCTGATGTGTGAGC 1380  
QY 461 ProLeuAspGlyGlnThrThrLysGlyLysThrLeuAspValArgAlaGlnAspGly 480  
Db 1381 CCTCTGACTATGACACAGACCAAGAGATACCTTACGGGTGCGAGCACAGATGATGCTC 1440  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1441 CGTCCCACTCTCTATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
Db 1501 AATGCCCTCATCTTGTGTGACAGCCCTTTCAGGGCTACTGCTCTGAGAGGCTCTCC 1560  
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1561 GGTACCTGCTTCTCATGCTCAGGCTATCAGCGTATGCTGATGCTGATGCTGCTGCTG 1620  
QY 541 GlnTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1621 GAATACCGCTTGTGCTGGGTGGGACATGACTTCCCTTCAACATGACATGACAGGC 1680  
QY 561 TrpIleSerValAlaAlaGlnLeuAspArgGlnGlnValAspPheThrSerPheGlyVal 580

```
Db 1681 TGGATCTCTGGCTGGTGAACCGGAGGAGTGTATTTCTACAGCTTGGGATA 1740
Qy 581 GIUALAARGSPHISGLYThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1741 GAAAGCTCAGAGACATGGGACCTCAGACACTGCTGGCGGCGAGTGTACGCTGACTCTC 1800
Qy 601 LeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db 1801 CTGGATGTCACAGAACATCCAACTTTACCCCAACAGATACAGAGTGGGCTCAAT 1860
Qy 621 GIUALAALAVALIYThrSerValIThrValSerAlaValAspAspAspAlaHis 640
Db 1861 GAGATGACGCTGTGGGACACAGCGGTGGTACGGGTGACGCTGTGACCGGTGATCTCAT 1920
Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db 1921 AGTGTCACTACCTTACAGATCAAGTGGCAATATCTCGAACCCTTCTCCATCCAGC 1980
Qy 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTylLeuGlnArgGln 680
Db 1981 CAAGTGTGGTGGCTGGTATCCCTTGCCTGCCACTGSACTGACAACTTGAGCGGAG 2040
Qy 681 TyrValLeuAlaValIThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db 2041 TATGTGTGGCTGTATACCGCTCCGATGACTGCGCAGAGACAGCAAGATGTGGTG 2100
Qy 701 AsnValIThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 2101 AATGTCAACGAGCCCAACCCATGCTCTGTCTTCAAGAGTCCCACTATACAGTGAAT 2160
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValIleuIleSerAlaIThrAspGluAsp 740
Db 2161 GTTATATGAGGACCGGCGCGGAGGACCAAGGTGGTGAATCAGCGGCAACGATAGGAC 2220
Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACGGTGGAATGGCCGATCACTTCACTTCAGTGAAGAACAGATCCCAAGTTCGCGATC 2280
Qy 761 AspAlaAspThrArgIleValIThrThrGlnIleGluLeuAspTyrGluAspGlnValSer 780
Db 2281 GATGACAGACAGGGGGCTGTCAACCAAGGCTGAGCTGAGCTAGAGACCAAGTGTCT 2340
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db 2341 TACACCTCGGCACTTACGTCTCGGACATGGCATTCCTCCAGAACTCCGACCTTAC 2400
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db 2401 CTGGAGATCTGTGGTGAACGCTGATATGACAAATGCCCTCAGTTCTCGAGACTCCCTAC 2460
Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGGCACTGTCTTGAAGAGATGTGCACCTTCACTAAGGTCTGCAATCTCAGCCACT 2520
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly 860
Db 2521 GATCGTGAATCTGGACTTAAATGGCGGCTTCTTCACTTCCAGAGGCGGACGATGGA 2580
Qy 861 AspGlyAspPheIleValIleGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GACGGTGACTTATATGTGAGTCCAGTCAAGGACATCGGCAACCTTACGAGAGCTGAT 2640
Qy 881 ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspGlyIleMetProPro 900
Db 2641 CGAGAGACGTTGGCCAGTATGTCTTGGGGGCAATGACAGTGAACAAGGAGATGCCCA 2700
Qy 901 AlaArgThrProMetGluValIThrValIThrValIleuAspValAsnAspAsnProProVal 920
Db 2701 GCCCGCACCACTTATGGAAGTGAAGTCACTGTGTGATGTGAAGAACATCCCTCTCTC 2760
Qy 921 PheGluGlnAspGluPheAspValPheValIleGluIleuAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTTGACAGATGAGCTTGTATGTCTTTGTGSAAGAGAACACCCCACTTGGGCTAGCCGTG 2820
Qy 941 AlaArgValIThrAlaThrAspProAspGluGlyIleThrAsnAlaGlnIleMetTyrGlnIle 960
Db 2821 GCCCGGTCAAGACATGACCCCGGATGAAGGACCAATGCCAGATTATATACAGATT 2880
Qy 961 ValGluGlyAsnIleProGluValIlePheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 2881 GTGAGGGCAACATCCCTGAGGTCTTCCAGCTGACATCTTCTCGGGGAGACTGACAGCC 2940
Qy 981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGTGATCTTGAAGTACAGAGACCGGCTGATGACGCTGTGTATCAGAGCCACGTGA 3000
Qy 1001 AlaProLeuValSerAlaValIleThrValIleValArgLeuLeuAspArgAsnAspPro 1020
Db 3001 GCTCCTCTGTGAGACCGGGGCTACAGTCCAGTCCGCTCTTGAACCGAATGACAACTCA 3060
Qy 1021 ProValIleuGlyAsnPheGluIleLeuPheAsnAsnTyrValIThrAsnArgSerSerSer 1040
Db 3061 CCAGTGTGGGCACTTGAAGATCTTTTCAACAACTATGTCAACAAATCGCTCAAGCAGC 3120
Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3121 TTCCCTGGGGGTGCATTTGGCGAGTACCTGCTCCATACCTGTATATCTCAGATAGTCTG 3180
Qy 1061 ThrTyrSerPheGluArgGlyAsnGlyLeuSerLeuValIleuLeuAsnAlaSerThrGly 1080
Db 3181 ACTTACAGCTTTGAGCGGGGAATGAATCAAGCTGTCTTCTCAATGCTCCACGGGT 3240
Qy 1081 GlnLeuIleSerAspArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal 1100
Db 3241 GAGCTGAAGCTAACCCCGGCACTGAGAACAAACCGGCTCTGAGGCGCATCATAGCGTG 3300
Qy 1101 LeuValSerAspGlyValIleSerValIThrIleGlnGlyValAlaValIleThrIleIle 1120
Db 3301 CTGTGTGACAGCGGCTGACAGGTACAGGCTACCGCAAGTGGCGGTGCGGTGATCACTATC 3360
Qy 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140
Db 3361 ACCGATAGAGATGCTCACCAAGCATCAGCTGTGGCTGAGAGACATGTACCCGAGCGC 3420
Qy 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaIleThrPro 1160
Db 3421 TTCTGTGACCACTGCTAGGCTCTTCACTCAGCGGTGGCGGCGGCAAGCGCA 3480
Qy 1161 ProAspHisValValIlePheAsnValIleArgAspThrAspAlaProGlyIleGlyIleIle 1180
Db 3481 CCGGACCACTGTGTGTCTTCAACGTACAGCGGACCAACGCGCCCGGGGGCCACATC 3540
Qy 1181 LeuAsnValSerLeuSerValIleGlnProProGlyProGlyGlyIleProProPheLeu 1200
Db 3541 CTCAACGTGAGCGTGTGGTGGGCGACCGCGAGGGCCGGGGGCGGCGCTTCTCTG 3600
Qy 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220
Db 3601 CCCCTTGAAGACCTGCGAGGCGGCTTATCTTCAACCGCAGCCTGTGACGCGCATCTCG 3660
Qy 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240
Db 3661 GCACAGCGGTGCTGCGCTTCAAGACCAACATCTGCTGGGAGAGCCCTCGAAGACTAC 3720
Qy 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260
Db 3721 ATGGCTGCGGTGTGGGTGCGCTTGAATCTTCCGCGGCTTCACTGCGCTCTCTCC 3780
Qy 1261 ValLeuPheArgProIleHisProValIleGlyIleuArgCysArgCysProProGlyPhe 1280
Db 3781 GTGTCTTTCGGGCGCATCCACCCGTCGAGAGGCGTGGCGGCTCCGCGGCGGCTTC 3840
Qy 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300
Db 3841 ACGGATACATACCTGCGAGACCGAGGTGACCTTCTGATCTCGCGGCGCTGTGGCCCCAC 3900
```



QY 1301 G1YArgCysArgSerArgGluGlyGlyTYrThrCysLeuCysArgAspGlyTYrThrGly 1320  
 DB GGGCGGCTGGCGGAGCGCGGAGGGGGCTTACACTGCTCTGTGTATGCTTACAGGGG 3960  
 QY 1321 G1uHicGysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIysAsnGly 1340  
 DB GAGCAGCTGTAGAGGTGAGTGTCTGCTCAGCGCTGTGCACTCCGGGTGTCTGCAAGAAATGGG 4020  
 QY 1341 G1YThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysProSerGlyAspPhe 1360  
 DB GGGACCTGTCTCAACTGCTGTGTGGGCGGCTTCAAGTGCATGTGCGCATCTGAGACTTC 4080  
 QY 1361 G1uIysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 DB GAGAGAGCCCTACTGCGCAGGTGACCAAGCGCAGCTTCCCGCCCACTCTTCACTCACTTT 4140  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIysGlyArg 1400  
 DB GCGGGCTTGGCAGCGCTTTCACCTTCACTGCGCTTCTGTGCGCAAGAGAGCGC 4200  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIysHisAspPheValAlaLeuGlu 1420  
 DB GAGGGGTGCTGTGTGTACATGGGCGCTTCAATGAGAGCATGACTTGTGGCCCTCGAG 4260  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
 DB GTGATCAGAGAGAGGTCCAGCTCACTTCTGCGAGGGAGTCAACACAGCGGTGTC 4320  
 QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuIleGlyTYr 1460  
 DB CCATTGTGTGCGGAGAGGTCACTGATGAGTGCAGTGTGATCGGTGCGCTGAATCTCTAC 4380  
 QY 1461 AsnIysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIysValAla 1480  
 DB AATTAACCACTGTGTGGTACAGAGGCTCCCAAGGCGCCCAAGAGCAAGAGGTGGCT 4440  
 QY 1481 ValValThrValIleAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB GTGTGATCAGCGGTGATGCTGTGACACAGAGTGGCTTGGCTTGGATGTGTCTGGGG 4500  
 QY 1501 AsnTyrSerCysAlaIleAsnGlyThrGlnGlyIleSerIysIleSerLeuAspLeuThr 1520  
 DB AACTTACTCCGTGTCTGCGGAGCGCACCGAGGTGGCGCAAGAAAGTCTGTGATCTGAGC 4560  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGlyIleSerPheProValArgMetArg 1540  
 DB GGGCCCTGTCTAAGCGCGGGGTGCTGACCTGCCAGAGCTTCCCAAGTCCCAATGCCG 4620  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 DB CAGTGTGTGGGTGACATGCGGAACCTGCAAGTGTGACAGCGGCACTATGACATGGCTGAC 4680  
 QY 1561 PheIleAlaIleAsnAsnGlyThrValProGlyCysProAlaIleIysIleAsnValCysAspSer 1580  
 DB TTCAATGGCCAAACATGTGACCGTGTGCTGCGCTGCCCAAGAAAGACGTGTGTGACAGC 4740  
 QY 1581 AsnThrCysHisAsnGlyGlyIleThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
 DB AACACTTGGCCACATGGGGGCACTGTGTGACCAAGTGGGACCGCTTCACTGCTGAGTGC 4800  
 QY 1601 ProLeuGlyPheGlyGlyIleIysSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
 DB CCCCTGGGCTTTGGGGGCAAGAGTGTGCGCCCAAGAAATGCGCAATCAGCAGCACTTCTCG 4860  
 QY 1621 GlySerSerLeuValAlaIleThrPheIleGlyLeuSerLeuProIleSerGlnProTyrIleu 1640  
 DB GGGAGCAGGCTGTGGGCTGTGCAATGGCTTCTGCTGCGCATCTTCCCACTGGTACCTC 4920  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValIleLeuGlnAlaIleThrArgGly 1660  
 DB AGCCTCATGTTCGCAAGCGCCAGGCGGACGCTGTCTGTGCAAGGCATCACAGGGGG 4980

QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 DB CGCAGACACATCACTCTACGCTTACAGAGGGCCACGTGATCTGAGCTGTGAGGGACA 5040  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 DB GGGCTTCAAGGCTCTCTCTCTGCTGTGAGCGCAGGCGCGGCAATGACGGTGATCTGGCAC 5100  
 QY 1701 HisAlaGlnLeuAlaLeuGlyValAsnGlyIleProGlyHisAlaIleLeuSerPheAsp 1720  
 DB CATGCAAGCTGTGCACTGGAGGCCAGGGGGGCGCTGTGCATGTCCATTTGTCTTCGAT 5160  
 QY 1721 TyrGlyGlnGlnArgAlaGlyGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleuSer 1740  
 DB TATGGGCAAGCAAGAGAGAGGGCAACTGGGCCCCCGGCTGCATGTGTGCACTGTGAC 5220  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
 DB AACATTAACATGTGGGGGGAATACCTGGGCGAGCGGCTGTGGCCCGCTTGTGGGGC 5280  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 DB TGTGTGCAAGGTGTGTGGGTGAGCCATACGCCAGAGGGGTTTAACTGATCTTCAGC 5340  
 QY 1781 HisGlyIleSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB CATGGGAGAGCATCAAGGTGAGCAAGGCTGTACCTGTGCTGCACTTGTGTACCTAAAC 5400  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyrSerCysSerCysAsp 1820  
 DB CCGTGTCTGTAAACGATATTTGACAGCAACATGAGCAAGTAACTTCTGCGAGCTGTGAT 5460  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAspProCysGlyHisGln 1840  
 DB CAGGTACTATGTAGCAACTGTACTATGTGTGTACCTGAACTCCGTGTAGACACAG 5520  
 QY 1841 SerValCysThrArgIysProSerAlaProHisGlyTYrThrCysGluCysAspProAsn 1860  
 DB TCTGTGTGTACCCGCAAGCCCGAGTCCCGCATGTGCTATACCTGTGAGAGTCCCAAT 5580  
 QY 1861 TyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysAspArgGlyTyrThrGly 1880  
 DB TACTTGGGCGCACTGTGTGAGCCAGATGTGACAGCGCTTGTCCCGGTGGCTGGGGGA 5640  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerIysGlyPheAspProAspCysAsn 1900  
 DB CATCCCAATGTGGGCCCATGTGCACTGTATGTGACCAAAAGCTTGTGACCCAGATCGAAC 5700  
 QY 1901 LysThrSerGlyGlyCysHisCysIysGluAsnHisIleTyrArgProProGlyIleSerProThr 1920  
 DB AAGACAAAGCGGGAGTGTGCACTGTCAAGAGAACTCACTACCGGCCCCAGGCGAGCCCCACC 5760  
 QY 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940  
 DB TGCCTTGTGTGATGTGTGATCCCAAGAGCTCTTGTGCAAGGTGTGTGACCTTAAGAT 5820  
 QY 1941 GlyIleCysProCysLeuPheProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 DB GGCAGGTGTGCATGTGACAGCGAGTGTCAATCGGGCGCTGAGTGTGACCGCTGTGACACCT 5880  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysAspArgAlaIle 1980  
 DB TTTGTGTGAGTGTACCAACATGTGTGAAAGTAAATTTATGACAGTGGCCCAAGACGGAAT 5940  
 QY 1981 GluAlaGlyIleTyrThrProArgThrArgPheGlyLeuProAlaAlaIleAspProCysPro 2000  
 DB GAGGCTGGGATTTGATGGGCCCGCTTCCGCTTCCGGCTGTGCTGTGCTCTCTGTCTCC 6000  
 QY 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGluHisArgGlyTyrLeuAspPro 2020  
 DB AAAGGCTCTTTTGGAGACTGTGTGCGCACTGTATGTAGACAGAGGGGGGTCTCCGCCCA 6060  
 QY 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuIysGlyPheAlaGluArgLeu 2040

Db 6061 AACCTTCAACATGACAGCTCCATCACTTCTCAGAACTGAAAGGCTTCGCTGAGCCGCTA 6120  
 QY 2041 GlnArgAsnGlyuSerGlyLeuAaspSerGlyAArgSerGlnGlnLeuLeuLeuAArg 2060  
 Db 6121 CAGCGAATACAGTCAAGGCTTCAAGCTCAGGAGGCTCCCAAGACCTTACGCTGCTTGGCC 6180  
 QY 2061 AenAlaThrGlnHisThrAlaGlyTyrPheGlySerAaspValylValAlaTyrGlnLeu 2080  
 Db 6181 AAGCGCACCGACAGCACAGCTGGCTACTTGGCAGACGACTCAAGGTGGCTTACAGCTG 6240  
 QY 2081 AlaThrArgLeuLeuAlaHisGlyuSerThrGlnAArgGlyPheGlyLeuSerAlaThrGln 2100  
 Db 6241 GCCACCGGCTGCTGGCCCAAGAGACCCAGCGGGCTTGGGCTGTCTGCAACACAG 6300  
 QY 2101 AaspValHisPheThrGlnAenLeuLeuAArgValGlySerAlaLeuAenAaspThrAlaAen 2120  
 Db 6301 GAGCTCACTTCACTAGAACTCTGCTCGGGGTGGGAGCGCTCTTGGACACAGCCAC 6360  
 QY 2121 LysArgHisTrpGlyuLeuLeuLeuGlnGlnThrGlnGlyTyrAlaTrpLeuLeuGlnHis 2140  
 Db 6361 AAGCGGCACTGGAGCTGATCCAGACAGAGGGTGGCACTGGCTGGCTGCCAGCAC 6420  
 QY 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAaspMetAArgHisThrTyrLeuSerProPhe 2160  
 Db 6421 TATGAGGCTTACCGCAGTGCCTGGCCCAAGACATCGGCACTTAAACCCCTTC 6480  
 QY 2161 ThrAlaValThrProAenAlaValHisSerValAlaArgLeuAaspGlyAenPheAla 2180  
 Db 6481 ACCATGTCACGCGCCCAACATTCCTCCGTAGTGCCTTGGACAAAGGAACTTGGCT 6540  
 QY 2181 GlyAlaLysLeuProAArgTyrGlnAlaLeuAArgGlyGlnGlnProAaspLeuGlnThr 2200  
 Db 6541 GGGGCCAAGCTGCCCGCTACAGAGCCCTGCGGGAGAGCGCCCGGACCTTGAGACA 6600  
 QY 2201 ThrValIleLeuProGlyuSerValPheAArgGlnThrProProValValAArgProAlaGly 2220  
 Db 6601 ACAATATTCCTGCTAGTCTGTCTTCAAGAGACCCCGCTGGTCAAGGCCCGGAGGC 6660  
 QY 2221 ProGlyGlnAlaGlnGlnProGlnGlnLeuAlaAArgGlnAArgArgHisProGlyuLeu 2240  
 Db 6661 CCGGAGAGGCCCCAGAGGCAAGAGAGCTGGCACGGGACAGGACCGGAGCTG 6720  
 QY 2241 SerGlnGlyGlnAlaValAlaSerValIleTyrAArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 6721 AGCAGAGGGTGAAGCTGTGGCAGCTCATCTACCGCACCTTGGCCCGGCTTACGCT 6780  
 QY 2261 HisAenTyrAaspProAaspLysAArgSerLeuAArgValProLysAArgProIleLeuAenThr 2280  
 Db 6781 CATTACTATGACCTCTACAGAGCGACGCTTGAGAGTCCCAAGCCCGATCATCAACACA 6840  
 QY 2281 ProValAlaSerIleSerValHisAaspAaspGlnGlnLeuLeuProAArgAlaLeuAaspLys 2300  
 Db 6841 CCGGTGTAGACATCAGCGCTCATGTATGAGAGACTTCTGCCCCGGGCCCTGGACAA 6900  
 QY 2301 ProValThrValGlnPheAArgLeuLeuGlnThrGlnGlnAArgThrLysProIleCysVal 2320  
 Db 6901 CCGGTACCGGTGACGCTTCCGCTGAGAGACAGAGAGCGGACCAAGCCCATCTGTGC 6960  
 QY 2321 PheTyrAenHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaAArgGlyCysGln 2340  
 Db 6961 TTCTGGAAACCATTAATCCGTGTCAGTGGACAGGTGGCTGTGGCCAGAGGCTGTGAA 7020  
 QY 2341 ValAlaPheAArgAaspGlnSerHisValSerCysGlnCysAenHisAmetThrSerPheAla 2360  
 Db 7021 GTGTCTTCCGCAATAGAGCCAGCTCACTGTCAGTGGCAACCATGACAGCTTCGCT 7080  
 QY 2361 ValLeuMetAaspValSerAArgGlyGlnAenGlyGlnIleLeuProLeuLysThrLeuThr 2380  
 Db 7081 GGCTCTATGACCTTCTCTCGGCGGAGAAATGGGGAATCTCTGCACCTGAAGCACTGACA 7140  
 QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400

Db 7141 TACGTGGCTTAGAGTGTCACTTGGCTGCCCTTGTGCTCACTTCTTCTTCACTTTC 7200  
 QY 2401 LeuAArgIleLeuAArgSerAenGlnHisGlyIleAArgAArgLeuThrAlaAlaLeuGly 2420  
 Db 7201 TTGGCTATCTGGCTTCCCAACACAGGACATCCAGATCCAGACTGACAGCTTGGCC 7260  
 QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAenGlnAlaAaspLeuProPheAlaCysThr 2440  
 Db 7261 CTGGCTCAGCTGGTCTTCTCTCGGGAATCAACAGAGGCTGACCTCCCTTGTGCTGACA 7320  
 QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGln 2460  
 Db 7321 GTCAATGGCATCTGTGTGACATTCCTTCACTTGCACCTTTCCTGGGCTGTGGAG 7380  
 QY 2461 AlaLeuHisLeuTyrAArgAlaLeuThrGlnValAArgAaspValAenThrGlyProMetAArg 2480  
 Db 7381 GCTTGGACCTTACCGGCACTCACTAGAGTGGCGGATGTACACCGGCCCCATGGCC 7440  
 QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
 Db 7441 TTCTACTACATGCTGGGCTGGGGCGTGGCTTCAATCAGAGGCTAGCCGTGGGCTG 7500  
 QY 2501 AaspProGlnGlyTyrGlyAenProAaspPheCysTrpLeuSerIleTyrAaspThrLeuIle 2520  
 Db 7501 GACCCGAGGGCTACGGGAACTTGACTGTGCTGGCTCTCATCTATGACAGCTCATC 7560  
 QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
 Db 7561 TGGAGTTTTCGTGGCCGGTGGCTTTGGCGCTGTGATGAGTCTTCTGTACATCTCG 7620  
 QY 2541 AlaAlaAArgAlaSerCysAlaAlaGlnAArgGlnGlyPheGlyuLysGlyProValSer 2560  
 Db 7621 GCGGCCCGGCTCTCTGTGTCTCCAGCGGCAAGGCTTGAAGAAAGGCTGTGCTCG 7680  
 QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
 Db 7681 GGCTTGCAGCCCTCTTGCCTTGCCTGTCTGTGAGGCGCACGTGGCTGTGGACGTG 7740  
 QY 2581 LeuSerValAaspSerAaspThrLeuLeuPheHisTyrLeuPheAlaThrCysAenCysIle 2600  
 Db 7741 CTCTGTGTCACAGGACACCTCTCTTCCATCTTGTGCTACCTTGTGCTGCAATGGCATC 7800  
 QY 2601 GlnGlyProPheIlePheLeuSerTyrValAlaLeuSerLysGlnValAArgLysAlaLeu 2620  
 Db 7801 CAGGCGCCCTTCATCTTCTCTCTATGTGGCTTAGACAGAGTCCGGAAGAACATC 7860  
 QY 2621 LysLeuAlaCysSerAArgLysProSerProAaspProAlaLeuThrThrLysSerThrLeu 2640  
 Db 7861 AAGCTTGGCTGAGCGGACCGGACCGCTTGAACCTGTCTGACCAACCAAGTCAACCTG 7920  
 QY 2641 ThrSerSerTyrAenCysProSerProTyrAlaAaspGlyAArgLeuTyrGlnProTyrGly 2660  
 Db 7921 ACCTGCTCTTCAACTGGCCCAAGCCCTTACGAGATGGGCGGCTTACAGCCGCTTACGA 7980  
 QY 2661 AaspSerAlaGlySerLeuHisSerThrSerAArgSerGlyLysSerGlnProSerTyrIle 2680  
 Db 7981 GACTCGCGCGGCTCTCTGACAGCACAGTCCCTCGGCGAGAGTCAAGCCGACTACATC 8040  
 QY 2681 ProPheLeuLeuAArgGlnGlnSerAlaLeuAenProGlyGlnGlyProProGlyLeuGly 2700  
 Db 8041 CCGTCTTGTGAGGAGAGAGTCCGACCTTGAACCTGTGGCAAGGGCCCTTGGCTGGGG 8100  
 QY 2701 AaspProGlySerLeuPheLeuGlnGlyGlnAaspGlnGlnHisAaspProAaspThrAaspSer 2720  
 Db 8101 GATCCAGGAGCGCTGTCTCTGGAAGTCAAGACAGGACATGATCTGACACGGAATCC 8160  
 QY 2721 AaspSerAaspLeuSerLeuGlnAaspAaspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
 Db 8161 GACAGTGAACCTGTCTTGAAGACCAACAGAGTGGCTCTTATGCTTACCACTCATCA 8220  
 QY 2741 AaspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 Db 8221 GACAGTGAAGAGAAAG 8280

QY 2761 TTPAASPserLeuLeuGlyProGlyValagluArgLeuProLeuHisSerThrProLyAsp 2780  
 DB 8281 TGGATATGCTGCTGGGGCCCTGGAGCAGAGAGACTGGCCCTGCACAGTACTCCCAAGAT 8340  
 QY 2781 GlyGlyProGlyProGlyValValProThrProGlyAspPheGlyThrThrAlaLysGlu 2800  
 DB 8341 GGGGGCCCAAGGGCCCTGGCAGAGGGCCCTGGCCAGAGACTTGGGAGACCAAGCAAAAGAG 8400  
 QY 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluValAsnGlyAlaLeuSerArg 2820  
 DB 8401 AGTATGGCAACGGAGGCCCTGAGAGAGGCTGGGAGAAATGAGATGCTTCTCGA 8460  
 QY 2821 GluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyLysGlyLys 2840  
 DB 8461 GAGGGGTCCCTAAGGCCCTTCCAGGGCTCTTCCGCCAGCTCCCAAAAGGATCTTAAAG 8520  
 QY 2841 LysLysCysLeuProThrThrLysSerGlyLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
 DB 8521 AAGAAATGTCTGCCACCATCAGCAGAGAAAGCAGCCCTCGCGGCTCCCTGGAGCAA 8580  
 QY 2861 CysThrGlySerSerArgGlySerSerAlaSerGlyLysSerArgGlyValProProPro 2880  
 DB 8581 TGCAAGAGGTCTTCCGGGGCTCTTCCGCTAGTGAAGGAGAGCCGGGGGCTCCCTGCC 8640  
 QY 2881 ArgProProArgGlnSerLeuGlnGluLeuAsnGlyValMetProLysAlaMet 2900  
 DB 8641 CGCCACCGGCCCGGAGAGCTCCAGAGAGCAGTGAACGGGGTCAATGCCATGCCATG 8700  
 QY 2901 SerLeuLeuValaglyThrValAspGluAspSerSerGlySerGluPheLeuPheAsn 2920  
 DB 8701 AGCATCAAGGACGACGCTGGATGAGAGACTCTCAGGCTCCGAATTTCTTCTTAAC 8760  
 QY 2921 PheLeuHis 2923  
 DB 8761 TTCTCTGCAAT 8769  
 RESULT 2  
 AX549238 10531 bp DNA Linear PAT 26-NOV-2002  
 LOCUS Sequence 523 from Patent WO02061087.  
 DEFINITION AX549238  
 ACCESSION AX549238  
 VERSION AX549238.1 GI:25813930  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 Burmer, G.C., Roush, C.T., and Brown, J.P.  
 Antigenic peptides, such as for G protein-coupled receptors  
 (GPCRs), antibodies thereto, and systems for identifying such  
 antigenic peptides  
 Patent: WO 02061087-A 523 08-AUG-2002;  
 JOURNAL Lifespan BioSciences, Inc. (US)  
 FEATURES  
 source 1..10531  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 2038 a 3353 c 2998 g 2142 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 10531  
 Score: 15545.00 Matches: 2923  
 Percent Similarity: 100.00% Conservativity: 0  
 Best Local Similarity: 100.00% Mismatch: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-916-849a-3 (1-2923) x AX549238 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20  
 DB 63 ATGGAGAGCCCGGCCACCGGGCTCCCCCTCCCAACCCCGCGCGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 DB 123 TTGCTGCTGCTGCTGCGCGCCACCATATGGGAGACCAAGTGGGGCTGCTGCTGCTG 182  
 QY 41 GlySerArgGlyValArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSer 60  
 DB 183 GGGTCCAGGGAGAGAGCTCTTTCGGGGCTGGGCCCAATGGGCTGCTGCTGCTGCTG 242  
 QY 61 SerAlaSerAsnLeuThrLeuThrLysSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 DB 243 TCAGGCTTCGAACTCTGGCTTACACACCGCTGAGGAGATGGGGCATGAGCTGACT 302  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuValValThrCysProGluSerGlyAlaHis 100  
 DB 303 GGCACCTGGTACCCCAACAGATGCTGAGGGGTTGGGTCTCAGAAATCCGAGGCCAT 362  
 QY 101 IleProLeuProProAlaProGluGlyCysProThrSerCysArgLeuLeuGlyLysGly 120  
 DB 363 ATTCCCTTACACAGACTCTTGAAGCTGCTGAGGCTGTGCTGCTGCTGCTGCTGCTG 422  
 QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGluGluHisAspProCysLeuValAla 140  
 DB 423 GGCACCTTTCCTCCCAAGGGCAAGCTCACATGCTCCGAGAGAGACCGGTAAAGGCT 482  
 QY 141 ProArgLeuArgCysGlnSerCysLysLeuValaglnAlaProGlyLysLeuValaglyGlu 160  
 DB 483 CCACGGCTCAGATGTCAGATCTGCAAGCTGGCAGCAGGCCCGGGCTCAGGGGAGGAA 542  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgLysArgAsnValAsnThrAlaProGln 180  
 DB 543 AGGTACAGAGAGAGTCTTGGGTGGGCTCGAAGAAAGAAATGTAATACAGCCCCCAG 602  
 QY 181 PheGlnProProSerTyrglnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 DB 603 TTCCAGCCCCCAGCTTACAGAGCCACAGTGGCGAGAACCAAGCAGGACCCCTGTT 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyValaglyValArgLeuGluThrMet 220  
 DB 663 GCATCCCTTGAAGGCCATTCACCCGACGAGGGTGAAGGCTGACTGATGATCACAGT 722  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 DB 723 GATGCCCTTTGATAGCCGCTCCAAACAGTCTTCTCCCTGAGACCACTACTGATGCA 782  
 QY 241 ValThrThrAlaGluGluLeuAspArgGlyThrThrLysSerThrHisValPheArgValThr 260  
 DB 783 GTAACCAAGCAGGAGAGGAGCTGATGATGATGAGACCAAGAGCACCAGCTTCAAGGGTCA 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaAlaValThrLeuThrLysLeuValThr 280  
 DB 843 GCGCAGAGCACGACGATGCTCCGACAGAGTCTTGGCTTACCTCAACATCTTGTTACT 902  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrglySerProSerGlyValPhe 300  
 DB 903 GACACCAAGACATACATCCCTGTGTTCAAGCAGCAGAGTACAGAGAGGCTCAGGGAG 962  
 QY 301 AsnLeuGlnValGlyTyrglyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB 963 AACCTGAGAGTGGCTATAGGTGCTCACTGTCAAGGGCCACGAGATGTATCCCTCC 1022  
 QY 321 AsnAlaAsnIleLeuThrArgLeuLeuGluGlySerProSerGlyValPhe 340  
 DB 1023 AATGCCAATATCTGATCCGCTGCTGAGGGGCTGAGGGGAGGCCCTTAAAGTCTT 1082  
 QY 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 DB 1083 GAGATGACCTCGCTGAGGGTGAATCCGATCCCGTGGCTCTGTGATGGGAGAGGGT 1142  
 QY 361 GluSerTyrglnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380

```

Db      1143 GAATCTTACCAAGTGAAGGCAAGTACAGGCTCGGAGCCCGGCTCTCGAAGT 1202
QY      381 ThrThrAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400
Db      1203 ACCAGAGCGCGCTGTTTCTTCTGTCGAGGATGACATATATATGCCCCCGATTAACT 1262
QY      401 GuluYarGlyValValGlnValArgGluAspValThrProGluValProValLeuArg 420
Db      1263 GAAAGACGCTATGTGTGTCAGTGAGGAGATGTGATCTCAGGGGCCCAAGTACCCGA 1322
QY      421 ValThrAlaSerAspArgAspGlySerAspAlaValIleGlySerIleMetSer 440
Db      1323 GTCAAGCCTCGGATGAGCAAGAGGAGCAATGCGGTGTGACATATAGCATCAAGAGT 1382
QY      441 GlyAspAlaArgGlyGlnPheTyrlLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460
Db      1383 GGCAGAGCTCGGGGACAGTTTATCTGAGATGCCCAAGCTGGACCTTGAGATGCGTAGC 1442
QY      461 ProLeuAspTyrlGluThrThrTyrlGlyTyrlThrLeuArgValArgAlaGlnAspGlyGly 480
Db      1443 CCTCTTGACTATAGACGACCAAGAGATACCCCTACGGGTGCGAGCACAGATGGTGGC 1502
QY      481 ArgProProLeuSerAspValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db      1503 CGTCCCACTCTATATGCTCTGCTGTGTGACAGTACAGGTCTCTGATATCAACGAC 1562
QY      501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      1563 AATGCCCATCTTCTGTCAGCACCCCTTTCAGGCTACTGTCTGAGAGAGCCCTTAA 1622
QY      521 GlyTyrlLeuValLeuIleValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db      1623 GGTCTACTGTGTCTCATGTCCAGGCTATCCAGCTGATGCTGTGTACATATGCCCTCTG 1682
QY      541 GluTyrlArgLeuAlaGlyValGlyIleAspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1683 GAATACCGCTTGTGAGGAGGACATGACTCCCTTCAACATCAACATGACACAGGC 1742
QY      561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrlSerPheGlyVal 580
Db      1743 TGGATCTCTGTGCTGCTGACAGCCGGAGAAAGTTGATTTCTACAGCTTTGGGGTA 1802
QY      581 GluAlaArgAspPheIleGlyThrProAlaLeuThrAlaSerAlaSerValThrVal 600
Db      1803 GAGAGCTCGAGACCATGTGCACTCCAGCACTCACTCGGCGAGTCTCAGCGTCACTGTC 1862
QY      601 LeuAspValAspAspAsnAspProThrPheThrGlnProGluTyrlThrValArgLeuAsn 620
Db      1863 CTGGATGTCAACGACAAACATCCACCTTAACTCAACAGAGTACACAGTGCCTCAAT 1922
QY      621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaIle 640
Db      1923 GAGGATGACGCTGTGGGACACAGCGTGTGACGCTGTGACCGTATATGCTCAT 1982
QY      641 SerValIleThrTyrlGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1983 AGGTGATACCACTTACAGATCACACAGTGTGCAATATCTGAAACCGCTTCTCATCACAC 2042
QY      661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrlTyrlLeuGlnValArgGln 680
Db      2043 CAAGAGGTGTGTGGCTGATCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2102
QY      681 TyrlValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      2103 TATGTGTGGCTGTATCCGCTCGATGGACCTGGCAGAGCACCGCACAGATGTGGTG 2162
QY      701 AsnValThrAspAlaAsnThrIleAsnProValPheGlnSerSerIleTyrlThrValAsn 720
Db      2163 AATGTACCGACCGCAACCATCTGTCTTTCAGAGCTCCCACTATATCAGTGAAT 2222
QY      721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740

```

---

```

Db      2223 GTTAATGAGACCGGCGGACGACCAACGCTGTGCTGATACGGCCACGATAGAGAC 2282
QY      741 ThrGlyGluAsnAlaArgIleThrTyrlPheMetGluAspSerIleProGlnPheArgIle 760
Db      2283 ACGGTGAGAAATGCCCGATCACTTACTTATGAGAGACAGCATGCCCATTCGGCATC 2342
QY      761 AspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyrlGluAspIleValSer 780
Db      2343 GATGACAGACACGGGGGCTGTACCAACCAAGCTGAGTGTGATTAAGAGAACAGATGTCT 2402
QY      781 TyrlThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleTyrlSerAspThrThrTyrl 800
Db      2403 TACACCTGTGCTATTAATGCTCTGCGAGCAATGCAATTCCTCCAGAAATCCGACCACTTAC 2462
QY      801 LeuGluIleLeuValAspAspValAspAspAsnAlaProGlnPheLeuArgAspSerTyrl 820
Db      2463 CTGAGATCTCTGTGATAACGATGATGATGATGATGATGATGATGATGATGATGATGATG 2522
QY      821 GlnGlySerValTyrlGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2523 CAGGCGACGTCTATGAGAGATGTGCCACCTTCACTACGCTCTGACAGTCTCAGCACT 2582
QY      841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrlThrPheGlnGlyGlyAspAspGly 860
Db      2583 GATCGTGAATCTGACCTTAATGACAGGGTCTTCAACCTTCCAGAGAGCGACATGGA 2642
QY      861 AspGlyAspPheIleValGlySerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db      2643 GACGCTGACTTATATTTGATGATCCACGTCAGGATCTGTGCGAACCTTACGAGGCTGAT 2702
QY      881 ArgGluAsnValAlaGlnTyrlValLeuArgAlaTyrlAlaValAspGlyGlyMetProPro 900
Db      2703 CAGAGAAACGTGGCCCACTATATGTCTTGGGCGCATGTGACATGACAAAGGAGTGCCTCA 2762
QY      901 AlaArgThrProMetGlnValThrValThrValLeuAspValAspAspAspProProVal 920
Db      2763 GCCCGCACCTATGAAAGTACAGTCACTGTGTGATGTGATGATGATGATGATGATGATGAT 2822
QY      921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2823 TTTGAGCAGAGATGATTTGATGTGTTTGTGAGAGAGAACACGCCATTTGGGTACCGTGTG 2882
QY      941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrlGlnIle 960
Db      2883 GCCCGGCTCACAGCCACGACCCCGATGAAAGCACCAATGCGCATATATGATACAGATT 2942
QY      961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db      2943 GTGAGGGCAATCTCTGAGGCTTCCAGCTGACATCTTCTCCGGGAGCTGACAGCC 3002
QY      981 LeuValAspLeuAspTyrlGluAspArgProGluTyrlValIleValIleGlnAlaThrSer 1000
Db      3003 CTGGTGAACCTTATGACTAGAGAGACCGGCTGATGATGATGATGATGATGATGATGATGAT 3062
QY      1001 AlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAspAspAspPro 1020
Db      3063 GCTCCTCTGTGAGAGCGGCTTACAGTCCAGTCCGCTCTTGAACCGCAATGACAAACCA 3122
QY      1021 ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrlValThrAsnArgSerSer 1040
Db      3123 CCAAGTCTGGGCAATTTGATGATCTTTTCAACAATATGTCAACAATGCTCAAGAGCC 3182
QY      1041 PheProGlyGlyAlaIleGlyArgValProAlaIleAspProAspIleSerAspSerLeu 1060
Db      3183 TTCCCTGGGGGGGCAATTTGGCGAGTATCTGCCATATGACCTGATATCTCAGATGATGCTG 3242
QY      1061 ThrTyrlSerPheGluArgGlyAsnGluLeuSerLeuValIleLeuAsnAlaSerThrGly 1080
Db      3243 ACTTACAGCTTGTAGCGGGGAAATGATCACTGAGCTGTGCTCAAGAGCTTCCACCGGGT 3302
QY      1081 GluLeuTyrlSerSerArgAlaLeuAspAspAsnAspProLeuGluAlaIleMetSerVal 1100
Db      3303 GAGCTGAAGCTTACCGCGGCACTGACACAAACCGGCTCTGTGAGGCAATCATGAGCTGT 3362

```

QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
 Db 3363 CTGGGTGACAGCGGCTACAGAGCGTGACCGCCAGCGCGCTCGGTGACCACTAC 3422  
 QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
 Db 3423 ACCGATGAGATGCTCACCAAGCATCAAGCGCTCCGAGAGACATGTCACTCCGAGCGC 3482  
 QY 1141 PheLeuSerProLeuLeuGluYleuPheIleGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
 Db 3483 TTCCGTGACCACTGTCTAGGCTCTTCAATCCAGCGGTGGCGGCCACGCTGACCA 3542  
 QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGluYglHisIle 1180  
 Db 3543 CCGGACCAAGTGTGTCTTCAACGTACAGCGGACACCAAGCGCCCGGGGGCCACATC 3602  
 QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGluYProGluYglYProProPheLeu 1200  
 Db 3603 CTCAACGTGACCTGTCTGATGGCCAGCGCCGCGGGGGCGGCGCCCTCTCTCG 3662  
 QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
 Db 3663 CCTCTGAGACCTGTGACAGAGCGCTTATCTTCAACCGACGCTGTGACGGCCATCTCG 3722  
 QY 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240  
 Db 3723 GCAAGCGCGTGTCTCTTCCGACGACCAATCTGCGCGGAGCGCCCTGGAGAACATC 3782  
 QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
 Db 3783 ATGGCTGCGGTGTGCTGTCTGCTGCTTCACTCTCCGCGCCCTTCACTGCTCTCTCC 3842  
 QY 1261 ValLeuPheArgProIleHisProValGlyGluLeuArgCysArgCysProProGluYpHe 1280  
 Db 3843 GTGCTCTTCCGGCCATTCACCCCGTCCGAGGGGCTGGCTGCGCTGCCCGCGGCTTC 3902  
 QY 1281 ThrGluAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3903 ACGGCTGACTACTGAGAACCGAGGTGACCTGTCTACTCGCGGCGCTGTGGCCCGCAC 3962  
 QY 1301 GlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysValArgAspGlyTyrThrGly 1320  
 Db 3963 GGGGCGCTGCGGACCGCGAGGCGGCTACACTGCTCTGTCTGTGAGGCTTACACGGGT 4022  
 QY 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGluYValCysLeuAsnGly 1340  
 Db 4023 GAGCACTGTGAGTGAGTGCTGCTGCTCAGGCGCTTGACCCCGGGTGTGTGAGAAATGG 4082  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheIleCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACCTGTGTCAACTGCTGTGGCGGTTTCAAGTGCATTCGCCCATCTGAGACTTC 4142  
 QY 1361 GluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 Db 4143 GAGAAAGCCCTACTGCTGACAGTACCAAGCGAGCTTCCCGGCCACTCTTATACACTTT 4202  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGluArg 1400  
 Db 4203 CGCGGCGTGGCGGCGTTCCTTCACTTCACTGCGCTCTCTGTTGGCCAAAGAGCGC 4262  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyValArgPheAsnGluYHisAspPheValAlaLeuGlu 1420  
 Db 4263 GACGGGTGTGTGTATCAATGGGCTTCAATGAGAGCATGACTTGTGTGGCCCTCGAG 4322  
 QY 1421 ValIleGlnGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440  
 Db 4323 GTGATCCAGAGACAGGTCAAGCTCACTTCTGTGAGGAGAGTCAACCAACGAGTGTCC 4382  
 QY 1441 ProPheValProGluYglValSerAspGlyGlnThrHisThrValGlnLeuLeuYserTyr 1460  
 Db 4383 CCAATTCGTGCGCGGAGAGTCAAGTGTGCGGAGTGTGATGAGTGTGAGTGTGAAATCTAC 4442

QY 1461 AsnLysProLeuLeuGlnGlnThrGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
 Db 4443 AATAGCCACTGTGGGTGACAGAGGCTCCCAAGGCGCCCAACAGAGAAAGTGGCT 4502  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 Db 4503 GTGTGTACCGTGTGATGTGTGTGACAGAGAGTGGCTTGGCGTTCGGATGTCTGTGGC 4562  
 QY 1501 AsnTyrSerValAlaGlnGlnYThrGlnGlySerTyrValSerSerLeuAspLeuThr 1520  
 Db 4563 AACTACTCTGTGTGTGCGAGGACCCAGAGTGGAGAGAAAGTCTGTGAATCTGACG 4622  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
 Db 4623 GGGCGCTGTCTTACGGCGGGGTGGCTGTGACTTCCCGAGAGCTTCCAGTCCGAAATGGC 4682  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 Db 4683 CAGTTCGTGGCTGTGACCGGAACTGTGAGTGAACAGCGCGCACATGACATGTGCTGAC 4742  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
 Db 4743 TTCAATGCCAAATGACACGTGTGCTGTGGCTGTGCTGCGCAAGAAAGAGTGTGTGACG 4802  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
 Db 4803 AACACTTGCACAAATGGGGGCACTTGCGTGAACCAATGGGAGCGCTTCAAGCTGCGAGTGC 4862  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
 Db 4863 CCCCCTGGCTTGGGGGCAAGACTGCGCCAGAGAAATGGCCAAATCAACAGCACTTCTCG 4922  
 QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640  
 Db 4923 GCGAGACGCTGTGTGCTGTGACATGTGCTCTGCTGCTCCATCTCCAACTGTGTACTTC 4982  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 4983 AGCCTCATGTTCGCGACGCGCCAGCGCGGAGGTGTCTGCTGAGGCCATCAACAGGGGG 5042  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlnHisValMetLeuSerValGluGlyThr 1680  
 Db 5043 CGAGACCACTCACTCACTGACGACAGAGGCGCAAGTATGTGTGAGCGTGGAGGCGACA 5102  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 Db 5103 GGGCTTCAAGGCTCTCTCTCTCTCTGTGAGCCAGGCGGGCCAAATACGTTGACTGGCAC 5162  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 5163 CATGCAACGCTGCACTGGGAGCCAGCGGGGGGCTGGCCATTCATTCGTGCTTCGAT 5222  
 QY 1721 TyrGlyGlnGlnArgAlaGluGluAsnLeuGluYProArgLeuHisGlyLeuHisLeuSer 1740  
 Db 5223 TATGGGAGAGAGACAGAGGAGGCAACTGGGCGCCCGGCTGTATGTCTGCACTGAGG 5282  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
 Db 5283 AACATAACAGTGGCGGGAATCACTGGGCGAGCGGGGTGTGGCCGTGGCTTCCGGGGC 5342  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 Db 5343 TGTTTGACAGGTGTGTGGGTGAGCGATAGCGCAGAGGGGGTAAACGCTGTGATCCACG 5402  
 QY 1781 HisGlyGlnSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 Db 5403 CATGGGAGAGCAATCAAGTGAAGAGGCTGTGAGCTGTGCTTACCTGTGTGCTCAAC 5462  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAsp 1820  
 Db 5463 CCGGTCTGTCTAAAGCTATTGTGACCAACGACTGGGAGACGTAATCTGCTGAGCTGTGAT 5522  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840

```
Db 5523 CAGGTTACTATGATGACAACTGACTAATGCTGACCTGAACCCGGTAGACACCG 5582
Qy 1841 SerValCysThrArgIysPheProSerAlaProHisGlyTyrThrCysGluCysProProAla 1860
Db 5583 TCTGTGTGATACCGGACGAGCCAGTGCCTCCCAATGGCTAATACCTGGGAGAGTGTCCCAAAAT 5642
Qy 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyTyrTrpGly 1880
Db 5643 TACCTTGGGCAATACCTGATGACCAAGAAATTGACCAAGCTTGTCCCGTGGCTGTGGGGA 5702
Qy 1881 HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn 1900
Db 5703 CATCCCAATGATGGCCCATGCAACTGTGATGTCAGAAAGCTTTGACCCCAACCTGCAAC 5762
Qy 1901 LysThrSerGlyGluCysHisCysGlyAsnHisIleTyrArgProProGlySerProThr 1920
Db 5763 AAGACAAAGGGGAGGAGCCACTGCAAGAGAACCACTACCGGCCCCCAAGACGCCCAAC 5822
Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp 1940
Db 5823 TCCCTCTGTGTGACTGCTACCCCAAGGCTCTGTCCAGAGTCTGTGACCTTGAGAT 5882
Qy 1941 GlyGlnCysProCysIleProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960
Db 5883 GGCACAGTGTCCATGCAAGCAGGTGTCATCGGAGCGTCAGTGTGACCGTGTGACAACTCT 5942
Qy 1961 PheAlaGluValThrThrAsnGlyCysGlyValAsnTyrAspSerCysProArgAlaIle 1980
Db 5943 TTTGTGTAGGTGACCAACCAATGGCTGTGAAAGTAATGACAGCTGCCCAAGAGCGATT 6002
Qy 1981 GluAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro 2000
Db 6003 GAGGCTGGGATCTGGTGGCCCGCTACCCGCTTCGGGCTGCTGCTGCTCTGTCCTCC 6062
Qy 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArgGlyTyrLeuProPro 2020
Db 6063 AAGGCTCTTTGGAGCTGTGGTGGCCACTGTGATGACAGCAAGGAGGTGGTCCCCCA 6122
Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuGlyPheAlaGluArgLeu 2040
Db 6123 AACCTTTCACTGACGATCGTCACTACCTTCTCAGAACTGAAGGCTTCGTGAGGGCTGA 6182
Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060
Db 6183 CAGCGGAATGAGTCAAGGCTTACGACTCAAGGCGCTCCCAAGCACTGCTCTCTGGGCG 6242
Qy 2061 AsnAlaThrGlnHisThrAlaGlyTyrArgPheGlySerAspValLysValAlaTyrGlnLeu 2080
Db 6243 AAGCGCAAGCAACACAGCTGCTACTTCGGCAGCGACGTCAGAAAGTGGCTTACCAAGCTG 6302
Qy 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100
Db 6303 GCCAAGCGGCTGCTGGCCCAAGAGAACCCAGCGGGGCTTTGGGCTGTCTCCACAGAC 6362
Qy 2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120
Db 6363 GACGCGCACTTCACTGAGAACTGCTGCTGGCGGTGGCAGCGCCCTCTGACACACCAAC 6422
Qy 2121 LysAspGlnIleTyrGlyLeuIleGlnGlnThrArgIleGlyTyrAlaIleThrLeuGlnHis 2140
Db 6423 AAGCGCACTGGGAGCTGATCCAGAGACAGAGGGGTGGCAGCGCTGGCTGTCCAGCAC 6482
Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisIleThrLeuSerProPhe 2160
Db 6483 TATGAGGCTTACGCAAGTGCCTGGCCCAAGAACAGCGGSCAACACTTACCTAAGGCCCTTC 6542
Qy 2161 ThrIleValThrProAsnIleValIleSerValValAlaGluAspLysGlyAsnPheAla 2180
Db 6543 ACCATCGTCAACGCCCAATGTGATCTCCGTAGTGGCTGTGACAAAGGGAATCTTCTCT 6602
Qy 2181 GlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlyGlnProProAspLeuGlnThr 2200
Db 6603 GGGGCAAGCTGCCCGCTACAGGCGCTGCGTGGGAGACAGCCCCCGGACCTTGAGACA 6662
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProValValArgProAlaGly 2220
Db 6663 ACAAGTATTTGCTGAGATCTGTCTTACAGAGAGCGCCCGGTGTGTACAGGCCCGCAG 6722
Qy 2221 ProGlyGluAlaGlnGlnProGluGluLeuAlaArgArgGlnArgArgHisProGluLeu 2240
Db 6723 CCGGAGAGGCGCCAGAGACCAAGAGAGCTGGCACCGGCAAGCGACCGGAGCGCTG 6782
Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260
Db 6783 AGCCAGGTTGAGGCTGTGGCCAGCTCATCATCTACCGCACCTTGCGGCTACTGCTCT 6842
Qy 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280
Db 6843 CATACATATGACCTTGACAAAGCGAGCTTGAGAGTCCCAAAAGCCGATCATCAACACA 6902
Qy 2281 ProValValSerIleSerValHisAspAspGluLeuLeuProArgAlaLeuAspLys 2300
Db 6903 CCGGTGTGACATCAGCGCTCATGATGATGAGAGCTTCTGCCCCGGGCGCTGACAAA 6962
Qy 2301 ProValThrValGlnPheArgLeuGlnThrGlnGlyArgThrIleProIleCysVal 2320
Db 6963 CCGGTCAAGTGTCCCTGCTGTGAGACAGAGAGCGAGCCAAAGCCATCTGTCTC 7022
Qy 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGlu 2340
Db 7023 TTTGTGAAACATTCATCTGTGTGATGAGTGCACAGTGTGCTGTGGCCAGAGCGCTGAA 7082
Qy 2341 ValValPheArgAsnGlySerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360
Db 7083 GTGCTCTTCCCAATGAGAGGCAAGTCAAGTGCAGGTCGCAACATGACAGAGCTTCCT 7142
Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGlyIleLeuProLeuIleValThrLeuThr 2380
Db 7143 GTGCTAGAGAGTTCCTCGCGGAGATGGGAGATCTGGCCACTGAAGACATGACA 7202
Qy 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400
Db 7203 TACGTGGCTAAGGTGATCACTTGGCTGCCCTTGTGTGATCTTCTTCTCTGACTCTC 7262
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420
Db 7263 TTGGGTATCCGCGCTCCAAACACAGGCACTCCGACGTACAGCAGCTGGCGCTGGGCG 7322
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440
Db 7323 CTGGCTAGCTGTGTCTCTCTGGAATCAACAGGCTGACTCTCTTGTGCTGACACA 7382
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuGln 2460
Db 7383 GTCAATTCGCACTCGTGCACCTTCTGTACTCTGCACTTTTCTGGGCTCTGCTGGAG 7442
Qy 2461 AlaLeuHisIleuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480
Db 7443 GCGTTGCACTGTGACCGGCACTCACTGAGGTGGCAGATGTCACACCGGCCCCATGGCG 7502
Qy 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValAlaGlyLeu 2500
Db 7503 TTCTACTACATGCTGGGCTGGGCGGTGCTGCTTCAACAGAGGCTGACCGGTGGCGCTG 7562
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520
Db 7563 GACCCCGAGGGCTACGGGAACCTTGACTTGTGCTGCTTCCATCTATGACAGCTCATC 7622
Qy 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540
Db 7623 TGAAGTTTGTGGCCCGGAGGCTTTGGCGGTTCAGATGATGTCTTCCTGTATCATCTCTG 7682
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnIlePheGlyLysGlyProValSer 2560
Db 7683 GCGGCGCGGCGCTCTGTGCTGCCACGCGGCGGCTTTTGAAGAAAGTCTGTCTCTG 7742
```



QY	2561	G1yLeuGlnIProSerPheAla1ValIleuLeuLeuLeuSerAla1aThrTPLeuLeuAlaIleu	2560
Dp	7743	GGCTCGAGCCCTCTTCGCGGCTCTCTGCTGCTGAGCGCAGTGGCTCTGGCACTG	7802
QY	2581	LeuSerValaAnSerAapThrIleuLeuPheh1s7YrIleuPheAla1aThrCYaAnCYa1Ile	2600
Dp	7803	CTCTCTGTCAACAGCAGACACCCCTCTTCCACTACCTCTTGGCACTCGAAATTGCACTC	7862
QY	2601	GlnG1yProPhe1IlePheLeuSerTyValaIleuSerTyVsg1ValaArgTyAlaIleu	2620
Dp	7863	CAGGAGCCCTTCATCTTCTCTCTCAATGGAGCTTACGAAAGAGAGTCCGAAAGCACTC	7922
QY	2621	LyLeuAla1aCYSerAArgLyProSerProAapProAlaIleuThrThyIleSerThyLeu	2640
Dp	7923	AAGCTTGGCTTCAGCCGCGAGCCCAAGCCCTGACCTGTCTGACCAACAAATCCACCTG	7982
QY	2641	ThySerSerTyAAnCYProSerProTyAlaAapG1ValaGLeuTyGlnProTyG1y	2660
Dp	7983	ACCTGTCCTTCAACAATGCCCCAGCCCTCAACCAATGGCGCGTTCACACCCCTTACGGA	8042
QY	2661	AapSerAlaG1ySerLeuH1SerThySerTyGserG1yLySerGlnProSerTyIle	2680
Dp	8043	GACTCGGCGGCTCTCTGCAACAGACCAAGTGGCTCGGGCAAGATCAAGCCCACTACATC	8102
QY	2681	ProPheLeuLeuAArgG1uSerAlaLeuAnProG1yGlnG1yProProG1yLeuG1y	2700
Dp	8103	CCCTTCTTGGCTGAGGAGAGATCCGCACTGAACCTGGCCAAAGGCCCCCTGGCCTGGGG	8162
QY	2701	AapProG1ySerLeuPheLeuG1uG1yGlnAapGlnGlnH1sAapProAapThrAapSer	2720
Dp	8163	GATCCAGACAGCCCTGTCTCTGAAAGTCCAAACCAAGCAAGATGATCTCTGAACGAGATCC	8222
QY	2721	AapSerAapLeuSerLeuG1uAapAapGlnSerG1ySerTyAlaSerThyIleSerSer	2740
Dp	8223	GACAGTACCTGTCTTAAAGAGACCAAGATGGCTCTATGGCTTACCCACTACATCA	8282
QY	2741	AapSerG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uAla1aPheProG1yG1uG1uG1y	2760
Dp	8283	GACAGTAAAGGAGAAAGAAAGAGAGAGAGAGAGGCGCTTCCCTGGAGACAGAGGC	8342
QY	2761	TryAapSerLeuLeuG1yProG1yAlaG1uAArgLeuProLeuH1sSerThyProLyAap	2780
Dp	8343	TGGATATAGCTCTGCTGGGGCTTGAGACAGAGATGACCTCCCTGCAACATCTCCAAAGAT	8402
QY	2781	G1yG1yProG1yProG1yLyVala1aProTyProG1yAapPheG1yThyThrAlaLyG1u	2800
Dp	8403	GGGGGCCCAAGGGCTTGGCAAGCCCCCTGGCCAGGAGATTTGGGACCAACAGCAAAAGAG	8462
QY	2801	SerSerG1yAAnG1yAla1aProG1uG1uAArgLeuAArgLyAAnG1yAAnLeuSerAArg	2820
Dp	8463	AGTATGTGCACACGGAGCCCTTGAGAGAGGCTGGGAGAAATGAGAGATCCCTGTCTCGA	8522
QY	2821	G1uG1ySerLeuG1yProLeuProG1ySerSerAlaGlnProH1sLyG1yI1eLeuLyS	2840
Dp	8523	GAGGGGTTCCTAGGCCCCCTTCCAGGCTCTTCTGCCAGGCTTCAAAAGGATCCTTAAG	8582
QY	2841	LyAlaLyCYLeuProThyIleSerG1uLySerSerLeuLeuAArgLeuProLeuG1uGln	2860
Dp	8583	AAGAAGTGTCTGCCCAACCATCAGCAGAGAAAGACGCTTCGGGCTCCCCCTGGAGCAA	8642
QY	2861	CyThyG1ySerSerAArgG1ySerSerAlaSerG1uG1ySerAArgG1yG1yProProPro	2880
Dp	8643	TGCACAGGATTTCCCGGGGCTCTTCCGCTATGTAGGGCAACCGGGGGGCCCCCTTCC	8702
QY	2881	ArgProProAArgInSerLeuGlnG1uInLeuAAnG1yValaMetProI1aAlaMet	2900
Dp	8703	CGCCCAACGCCCCGGCAGAGCTTCCAGAGACGCTGAACGGGGATCATGCCATCGCACTG	8762
QY	2901	SerI1eLyAlaG1yThyValaAapG1uAapSerSerG1ySerG1uPheLeuPheAAn	2920
Dp	8763	AGCATCAAGGAGGACGATGATAGGATCTGTCAAGGCTCCGAATTTCTTTCTTTAAC	8822

[illegible]



SRHIDMADPFIANNGTVCPCPAKKNVCDSTNCHNGTCVNDMDA95CECPDIFGQKSCA  
QEMANPDRHSSLVAMHGLSLPISQPMYLSIMFRTQADVLLOAIIRGSTTLOL  
REGVMHLSVEGTGLQASSLRLEPRGANGDMHMOALAGASGPGHAILLSDYQORA  
EKNLGPRLHGLHLSNITVYGIPGPAAGVPRGGLTQGVRSVDPBGNNLPSHSES  
INVEGCSLPDPCDNPCCPANSYCSNDMSYSCSDPCRYGDNCTNVCNLIAPCHQSV  
CTRKSAHGHGTCCEPNTLPGYCESTRIDQPCPRGMWHPTGCPNCVSKSPFDON  
KTSQCHCKENHMRPSPSPCLCDYCPYGLSRVCDDEDCQCPKCPVIRQDCRCD  
NPFATVNGCEVNDSCPRALENGIWMPRTFGLPAPAPCPKSPFGTAVHCDHRG  
WLPPLFNCTSTISFELKGFARLQDNSSGLDSRQOALLRVAOTHTYGFESDV  
KVAOLATRLAHSSTORPGLSATODPHNTLRSVLSALIDTNKHMELI00TEG  
GTAMLOHYEAVASALAQMRHTYLSPTTYTPNVIVISVVALDKENPRGALCPREAL  
RGSPDPLETYYILPESVFRPTPVVRPRRGAQSPREARHRELPBESLQGEVAVS  
VIIXKTLAGLPHANDPKRSLRVPKRIIMPVSVISHDEBELPALRKPVAVOF  
RLLEBETKPICVFMNHSILVSGTGSARGCEVFNESHSVSCQCHMTSPAVLMD  
VSRNENGILPLKTLTYVALGVTALALITPFLTLRLIRSNHGHIRNLTALAGLA  
OLVFLGINOADIPLACTVIALILHPLILCTPSPWALILALILYVALTEVRVNGPME  
FYVNLGKGVPAPIFGALVADLDEGNGDFCMLSTYDPLINSFAPVAPANYSMTFLY  
ILARASCAQORQERKGPVSGLOPSFAVILLSATWLLLSVNSDTLLPHYLFAT  
CNCIQGPPIPLSYVLSKEVRKALKLACSRKSPDPALTTSTLTSYNCSPPADGR  
LYOPYGDAGSLHSTRSKSPSYIPLTLREBSALNPGQPGIADPGSLFLBQDQ  
QHDPTDSDSLSDQSGYASTHSDBSEBEEBEEBAAPBQGMDSILGGAIE  
RLPLHSTKDGPGPGKAPWPGDPTGTAKESNGA PERRLRENDLSTRSGSLGPI  
GSSAQPHKGIKKKCLPTISKSSSLRPLEQCTSSSGSSASBSRGGPPPPPPRO  
SLQBLNCEVMPHMSIKAGIVDESDSSSEFLPFNPLH"

BASE COUNT 2038 a 3353 c 2998 g 2142 t

## ALIGNMENT Scores:

Pred. No.: 0 Length: 10531  
Score: 15545.00 Matches: 2923  
Percent Similarity: 100.00% Conserved: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x AF234887 (1-10531)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
DB 63 ATGGGAGACCCGGGACCCGGGCTCCCTCCCAACCGCGCGCGCTGCTGTG 122  
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyValGlnValGlyProCysArgSerLeu 40  
DB 123 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyTyrProLeuCysProSer 60  
DB 183 GGGTCCAGGGGACGAGGCTCTGCGGGGCTGCGCGCCCAATGGGCTGCTGCTCAATCC 242  
QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGlyLeuThr 80  
DB 243 TCAAGGTGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100  
DB 303 GGGCCACTGTGTAACCCACACGATGGCTGAGGGTGTGTGTGCTCCAGAAATCCGAGGCCAT 362  
QY 101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyTyrLeGly 120  
DB 363 ATTCCCTTACCAACGAGCTCTGAGAGGCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 422  
QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnAlaProCysLeuLeuValAla 140  
DB 423 GGGCCACTTTCCCAACGAGGACCTGACACTGCGCCGAGGACACCGTGTAAAGCT 482  
QY 141 ProArgLeuValGlyGlnSerCysGlyValLeuAlaGlnAlaProGlyValLeuArgValAla 160  
DB 483 CCACGGCTCAATGCTGCTGCAAGCTGCAACGGCCCCCGGGCTCAGGGCCAGGGGAA 542  
QY 161 ArgSerProGlnGlnSerLeuGlyValYArgArgValArgAsnValAsnThrAlaProGln 180  
DB 543 AGGTACCAAGAAAGTCCCTGGGTGGGCTCGAAGAAAGAAATGTAAATACAGCCCCCAG 602

QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnLeuGlnProAlaGlyThrProVal 200  
DB 603 TTCCAGACCCCACTACCAAGCCACAGTGCAGGAGAACCAAGCAAGGACCACTCCCTTT 662  
QY 201 AlaSerLeuAlaGlyAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlnTyrThrMet 220  
DB 663 GCATCCCTGAGGGGCACTACCAAGCCCAAGGAGGTAGGAGGCTCACTGAGTACCACTATG 722  
QY 221 AsnAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyVala 240  
DB 723 GATGCCCTCTTGAATAGCCGCTCCACCAAGTCTTCTCCGAGACCCAGTCACTGCTGCA 782  
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyLeuThrIleSerThrHisValPheArgValThr 260  
DB 783 GTAACCAAGCCGAGGAGCTGATCGTAGACCAAGAGCAACCACTTTCAGGGTCAAG 842  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrIleThrIleLeuValThr 280  
DB 843 GCGCAGAGCAACGCGCAATGCCCGCAAGAGTCCCTGCTACACTACCATCTGTGTTACT 902  
QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
DB 903 GACACCAATGACCAATGACCTGCTGTGAGAGCAAGAGTACCAAGAGGCTTCAAGGAG 962  
QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 963 AACCTGAGAGGTGGCTATGAGTGTCTCACTGACGGCCACGATGTGATGATGCCCTTCC 1022  
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340  
DB 1023 AATCCCAATATTCGTGACCCGCTGCTGAGGGGCTGAGGGGCAAGCTTGAAGTCTTT 1082  
QY 341 GlnIleAspProAspArgSerGlyValIleArgThrArgGlyProValAspArgGlnVal 360  
DB 1083 GAGATCCAGCTGCTGCTGAGGTATCCGAACCTGTGCTGAGTCCGAGGAGAGGTG 1142  
QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
DB 1143 GAATCTTACCAAGCTGAGAGTGAAGCAAGTGAACAGAGGTGGGACCGGGGCTCGAGGT 1202  
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAsnAspAlaProGlnPheSer 400  
DB 1203 ACCACACCGCTGTTTCTCTTCTGAGAGATGACATGATATATGCCCCCACTTAAGT 1262  
QY 401 GluIleAspArgTyrValValGlnValArgGlnAspValThrProGlyValaProValLeuArg 420  
DB 1263 GAGAGGCTATGTGTGCTGAGGTAGAGAGGATGTGATCCAGGGGCCCCAGTACTCGGA 1322  
QY 421 ValThrAlaSerAspAspArgAspGlySerAsnAlaValAlaHisTyrSerIleMetSer 440  
DB 1323 GTCAACCTCTCGATGAGACCAAGGGAGCAATGCCGTGTGCTATATGATCATGATGAT 1382  
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAspValValSer 460  
DB 1383 GGCATATCTGGGGACAGTGTATCTGATGCTGAGCTGGAAGCTCTGATGTGTGATGC 1442  
QY 461 ProLeuAspTyrGlnThrThrIleArgGlnTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
DB 1443 CTTTGAATATGAGACCAACAGAGTATACCTTACGGGTGCAAGCAACAGATGTGTGGC 1502  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 1503 CGTCCCACTCTCTATATGTCTGTGCTGTGTGATGATGATGATGATGATGATGATGATGATGAT 1562  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
DB 1563 AATGCCCACTCTTGTGACACACCTTTTCAGGACTATGCTCTGAGAGAGCTCCCTTA 1622  
QY 521 GlyTyrLeuValIleHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
DB 1623 GGTACTCTGTCTTCCATGCTCAGGCTATGACAGCTATGCTGATGATGATGATGATGATGATGATGAT 1682  
QY 541 GlnTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560

```

Db      1683 GAATACGGCTTGTGGGGTGGGACATGATCCCTTTCACATCAACATGGCACAGGC 1742
Qy      561 TPTLlSerValAlaIaGluLeuAspArgGluGluValAspPheTyrSerPheGluVal 560
Db      1743 TGGATCTCTGTGGCTGTCTGAATCGACCGGGAGGAGTTGATTTCTACAGCTTTGGGGTA 1802
Qy      581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      1803 GAAGCTCGAACAATGGCACTTCAGACCTCACTGCTCGGCAGGTGCAGCGTGAATGTC 1862
Qy      601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      1863 CTGGATGTCAACGACCAACATTCACCTTACCAACAGATGACAGTGGGGCTCAAT 1922
Qy      621 GluAspAlaAlaValGlyThrSerValAlaThrValSerAlaValAspArgAspAlaHis 640
Db      1923 GAGGATGCACTGTGGGACACAGGTGTGACGGTGTCACTGTGACCTGTGATGCTCAT 1982
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      1983 AGGTATATCACTACCAAGATCAACAGTGGCAATCTGAAACGGCTTCTCATCAACAGC 2042
Qy      661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIryLeuGluArgGln 680
Db      2043 CAAGTGTGTGGTGGCTGTATCCCTTGGCCCTGGCACTGACCTGACCAACTTGAGCGGAG 2102
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      2103 TATGTGTGGCTGTATCCGCTCCGATGGCACTCGGAGGACAGCGGACAGATTTGGTGT 2162
Qy      701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      2163 AATGTACCGACCGCAACCAACCATTCCTGTCTTCAAGCTCCCATCTACAGTGAAT 2222
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAsp 740
Db      2223 GYTAAGAGAGACCGGCGGAGGACCAACGAGTGTGTGATCAACCGGCAAGATGAGGAC 2282
Qy      741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2283 ACAGGTAGAAATGCCCATCACTACTTCAATGAGAGACAGATCCCGCATGTCGGATC 2342
Qy      761 AspAlaAspThrGlyAlaValIleThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2343 GATGACAGACACGGGGGCTGTCAACACCAAGCTGAGTGAAGTACAGAAACCAAGTCTC 2402
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIrySerAspThrTyr 800
Db      2403 TACACCTGGCCATTTACTGCTCGGGAACATGGCAATCCCAAGAGTCCGACACACTTAC 2462
Qy      801 LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyr 820
Db      2463 CTGGAGATCTCTGTGAACAGATGAATGACATGCCCTCAAGTCTCTGCAAGACTCTTAC 2522
Qy      821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2523 CAGGGCAGTGTCTATGAGATGTGCCACCTTCACTAGCGTCCCTGAGATCTCAGCCACT 2582
Qy      841 AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyGlyAspArgGly 860
Db      2583 GATCGTATTTCTGACCTTAATGCGAGGGTCTTTCACACTTCAACAGAGAGCGACAGTGA 2642
Qy      861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880
Db      2643 GACGGTACTTTATTTGTAGTCAACGTCAAGCACTCGTGGAAACGTCAGGAGGCTGAT 2702
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIryGlyMetProPro 900
Db      2703 CGAGAGAACTGGGCCAGTATGTCTTGGGGCATATCAGTGGCAAGGGGAGATGCCCA 2762
Qy      901 AlaArgThrProMetGluValIleThrValThrValLeuAspValAsnAspAsnProProVal 920

```

---

```

Db      2763 GCCCGACACTTATGAGAAATGACAGTCACTGTGTTGATGTGAATGACAAATCCCTGTC 2822
Qy      921 PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2823 TTTGACAGAGATGATGTGATTTGTGGTGAAGAGAACAGCCCATTTGGGCTAGCCGTG 2882
Qy      941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      2883 GCCCGGTACAGCCATCACTGACCCCATGAAAGGACCAATGCCAAGTTATGTACAGATT 2942
Qy      961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThrAla 980
Db      2943 GTGAGGGGCAACATCCCTGAGGCTTTCAGCTGAGCATCTTCTCCGGGAGCTGACAGCC 3002
Qy      981 LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db      3003 CTGGTGAACCTTGAACCTTACAGAGACCGGCTGAGTACGTCTGTGATTCAGAGCCAGTCA 3062
Qy      1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020
Db      3063 GCTCTCTGTGAGCGCGGCTACAGTCCAGTCCGCTCTTGAACCGCAATGACCAACCA 3122
Qy      3123 ProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db      1041 PheProGlyGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3183 TTCCCTGGGGGTGCCATTTGGCGAGTACCTGCCCATGACCTCGAATATCTCAATATGTCG 3242
Qy      1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      3243 ACTTACAGCTTTGAGCGGGGAATGACTCAAGCTGCTGCTCTCATATCCCTCCAGCGGT 3302
Qy      1081 GluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100
Db      3303 GAGCTGAAGCTTAAGCGCGCACTGGAACAACACCGGCTCTGAGGCGCATATGAGCGTG 3362
Qy      1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120
Db      3363 CTGGTGTCAAGCGGCTACAGCTGACCGCCCATGTCGCGCTGTGTACCATTCATC 3422
Qy      1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140
Db      3423 ACCGATGAATGCTCAACCAACGATCAACGCTGCGCTGAGAGACATGTCACCCGAGCGC 3482
Qy      1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160
Db      3483 TTCTGTGTCAACCTGTAGGCTCTTTCATCCAGGGGTGTGGCGCACAGCTGGCCACGGCA 3542
Qy      1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180
Db      3543 CCGGACACGTGTGTGTCTTCAACGTAACAGGGGACACCGAGCGCCCGGGGGGCCATTC 3602
Qy      1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyIleProProPheLeu 1200
Db      3603 CTCACGTGAAGCTGTGCTGTGGGCGAGCGCGGAGGCGCGGGGGCGGCGCTTCTCG 3662
Qy      1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIleSer 1220
Db      3663 CCTCTGAAGACCTTGAGAGAGCGCTTATACCTCAACCGGAGCTGTGACGGCAATCTCG 3722
Qy      1221 AlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCysGluAsnTyr 1240
Db      3723 GCACAGCGGTGCTGCTGCTTCAACAGCAACATCTGCTGGGAGAGCCCTGCAAGAACTAC 3782
Qy      1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260
Db      3783 ATGCGGTGTGTGTGCTGTGCTTCACTCTCGCGCCCTTCACTCCCTCTCTCC 3842
Qy      1261 ValLeuPheArgProIleHisArgProValGlyGlyLeuArgCysArgCysProProGlyPhe 1280
Db      3843 GTGCTCTTCCGGCCCATTCACCCCGTGGAGGGGTGGCTGGCGCTGCGCCCGGCTTC 3902

```

QY 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3903 ACGGTGACTACTCGAGACCGAGGTGGAACCTGCTACTCTGGGCGCTGTGGCCCCAC 3962  
 QY 1301 G1YArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 Db 3963 GGGGCGTCCGCGACCGCGAGGGCGGCTACCTGCTGTCTGTATGCTACACGGGT 4022  
 QY 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysValAsnGly 1340  
 Db 4023 GAGCACTGTAGAGTGAAGTCTCGCTCAGGCGGTGACCCGGGTGTCTGCAGAAATGAG 4082  
 QY 1341 G1YThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
 Db 4083 GGCACCTGTGTCAACCTGTGTGGGGGGTTCAAGTGCATTGCCATTCGAGGACTTC 4142  
 QY 1361 GluLeuProTyrCysGluValThrThrArgSerPheProAlaHisSerPhe1IethrPhe 1380  
 Db 4143 GAGAAAGCCCTAATGCCAGGTGACCAAGCGCAGCTTCCCGCCCATCTTCATCACTT 4202  
 QY 1381 ArgGlyLeuLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIysGluArg 1400  
 Db 4203 CGGGGCTTGCGCCAGCGCTTTCACCTTACCTTGCGCTCTGTGTGCAAGAGAGCGC 4262  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIshisAspPheValAlaLeuGlu 1420  
 Db 4263 GACGGGTGTGCTGTATCAATGGCGCTTCAATGAGAAAGCATGACTTGTGGCCCTCGAG 4322  
 QY 1421 ValIleGlnGluValGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
 Db 4323 GTGATTCAGAGAGCGGTGTCAGCTCACTTCTGCAAGGAGTCAACCAACGAGTGTCC 4382  
 QY 1441 ProPheValProGlyValSerAspGlyIshisThrValGlnLeuLeuTyrTyr 1460  
 Db 4383 CCAATTCGTGCGCGAGAGTCAAGTGAAGGCAAGTGGCAATGAGGACGCTGAATACTAC 4442  
 QY 1461 AsnLeuProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnValAla 1480  
 Db 4443 AATTAAGCAGCTGTGGGTCAAGACAGGGCTCCACAGGGCCCATCAAGAGAAAGTGGCT 4502  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 Db 4503 GTGTGTACCGTGTGATGCTGTGACACAGAGTGGCTTGTGATCTGTGCTGTGGGC 4562  
 QY 1501 AsnTyrSerCysAlaIleGlnGlyThrGlnGlyGlySerIshisSerLeuAspLeuThr 1520  
 Db 4563 AACTACTCTGTGTGCTGCCAGGGCAACCAAGGTGAGCAAGAAAGTCTGTGATCTGACG 4622  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
 Db 4623 GGGGCGCTGTACTAGCGGGGTGCTGACCTGCGGAGAGCTTCCAGTCCGAAATGCGG 4682  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 Db 4683 CAGTTCGTGGCTGTGACAGCGAACTGCAAGTGTGACGCGGCACTAAGACATGGCTGAC 4742  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaIshisValAsnValCysAspSer 1580  
 Db 4743 TTCAATTCCAACATGAGCAAGTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4802  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnIshisAspAlaPheSerCysGluCys 1600  
 Db 4803 AACACTTGCCACATGGGGGCACTGTGTACCAAGTGGGACGCGTTCACTGCGAGTGC 4862  
 QY 1601 ProLeuGlyPheGlyGlyIshisCysValGlnGlnIshisAsnProGlnIshisPheLeu 1620  
 Db 4863 CCCTGTGGCTTTGGGGCAAGAGTGTGCGCCAGAAATGCCCAATCAAGCACTTCTCG 4922  
 QY 1621 GlySerSerLeuValAlaIshisGlyLeuSerLeuProIshisGlnProTyrPheLeu 1640  
 Db 4923 GGCAGCAGCTGTGTGCTGTGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 4982

QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValIleLeuGlnAlaIleThrArgGly 1660  
 Db 4983 AGCTCATGTTCCGCAAGCCCGAGCGGACGAGTGTCTGTGTGAGGCTACCAAGGGG 5042  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 Db 5043 CGCAGACCATCAACCTTACAGCTTACAGAGGGGCAAGTATGCTGAGCGTGAAGGGCA 5102  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
 Db 5103 GGGCTTCAAGGCTTCTCTCTCTGCTGTGAGCCAGGCGGGCCAAATACGGTGACTGGAC 5162  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 5163 CATGCAACAGTGTGCACTGGAGGCAAGCGGGGCGCTGGCATTCCTGTCTTCGAT 5222  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 Db 5223 TATGGCAGCAGAGAACAGAGGGCAACTGGGCGCCGCGCTGCATGTGCACTGAGC 5282  
 QY 1741 AsnIleThrValGlyGlyIshisProGlyProAlaGlyValAlaAsnGlyPheArgGly 1760  
 Db 5283 AACATAAGTGGGCGAAATACCTGGGCGAGCGCGGTGTGCGCTTGTGGGGCT 5342  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 Db 5343 TGTTCACAGGTGTGTGGGTGAGCCATACCGCAAGGGGTTTACAGCTGTGATCCAC 5402  
 QY 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 Db 5403 CATGGGAGAGCATCAACGTGAGCAAGGCTGTAGCTCGACCTGTGTGCTCAAAC 5462  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyrSerCysSerCysAsp 1820  
 Db 5463 CGGTGTCTGTCAACGCTATTGTGACAGCAACAGCTGGAAAGCATATTCGCACTGTGAT 5522  
 QY 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluIshisGln 1840  
 Db 5523 CAGGTTACTAATGTGACAACTGTACTAATGTGTGATCTGAACCCGTTGTGAGCAACG 5582  
 QY 1841 SerValCysThrArgLeuAspSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 Db 5583 TCTGTGTGACCGCAAGCCAGGTGCGCCCATGTGTATACCTGTGCAAGTGTCCCCCAAT 5642  
 QY 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProArgGlyIshisProGly 1880  
 Db 5643 TACCTTGGGCAATCTGTGAGACAGGATTAACAGCTTGTCCCGTGTGTGTGGGGA 5702  
 QY 1881 HisProThrCysGlyProCysAsnCysAspValSerIshisGlyPheAspProAspCysAsn 1900  
 Db 5703 CATCCCAATGTGGCCAGTGCATGTGATGTCAAGAAAGGCTTTGACCCAGACTGCAAC 5762  
 QY 1901 IysThrSerGlyGluCysHisCysIshisGluAsnHisIshisArgProProGlySerProThr 1920  
 Db 5763 AAGACAAAGGGGAGAGTGCACAGCAAGGAAACCACTACCGGCCCCCAAGGAGCCCAAC 5822  
 QY 1921 CysLeuLeuCysAspCysThrProThrArgIshisSerLeuSerArgValCysAspProGluAsp 1940  
 Db 5823 TGTCTTGTGTGATCTGTACCCCAAGGCTCTTGTCCAGAGTCTGTATACCTGTGAGAT 5882  
 QY 1941 GlyGlnCysProCysLeuAspProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 Db 5883 GGCAGGTGTCCATGCAAGCAGGTGTCAATCGGGCGTCAAGTGTGACCGCTGTACAACT 5942  
 QY 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle 1980  
 Db 5943 TTGTGAGAGTCAACCAATGCGTGTGAAGTAATTAAGACAGCTGCCCAAGAGGAT 6002  
 QY 1981 GluAlaGlyIleTyrThrProArgThrArgPheGlyLeuProAlaAlaIshisProCysPro 2000  
 Db 6003 GAGGCTGGATCTGTGTGGGCGGCTTACCGGCTTGGGCTGTGTGCTGTGTGTGTGT 6062  
 QY 2001 IysGlySerPheGlyThrAlaValArgHisCysAspGlyHisArgGlyTyrLeuProPro 2020

Db	6063	AAAGGCTCTTTGGGACTGTGTGGCCACTGTGATGACACAGGGGGTGGCTCCCCCA	61222
Qy	2021	AsnLeuPheAncyEThrSerIleThrPheSerGluLeuYsgIyPheAlaGluArgLeu	2040
Db	6123	AAACCTTCACTGACGCTCCATCACTTCTCAGAACTGAAGGGGCTTGCTGAGGGCTTA	6182
Qy	2041	GlnArgArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg	2060
Db	6183	CAGCGGAATGAGTCAAGGGCTTAGACTCAAGGGGGCTCCAGCAGCTAAGCTTCTCTGGC	6242
Qy	2061	AsnAlaThrGlnHisIleThrAlaGlyTyPheGlySerAspValIysValAlaTyGlnLeu	2080
Db	6243	AAACGGACAGGACACACAGCTGTGCTACTTCGGCAGCGACGTCAAGGTGGCTTACCAAGT	6302
Qy	2081	AlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln	2100
Db	6303	GCCACGGCGCTGTGGCCCAAGAGAGCCACGGCGGGCTTGGGCTGTCTGCGCACAG	6362
Qy	2101	AspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
Db	6363	GACGTGCACCTTCACTGAGAAATCTGTGGGGGTGGGACGGCCCTCTTGACACAGCCAAAC	6422
Qy	2121	LysArgHisTrpGluLeuLeuIleGlnGlnThrGlnGlnGlyTyThrAlaTrpLeuGlnHis	2140
Db	6423	AAGCGGCACTGGAGACTGATTCAGACAGAGAGGGGTGGCACCGCTGTGCTCTCCAGCAC	6482
Qy	2141	TyrGlnAlaTyPAlaSerAlaLeuAlaGlnAsnMetArgHisEThrTyPLeuSerProPhe	2160
Db	6483	TATGAGGCTTACGCCAGATGAGCTCTGGCCCAAGACATGGGACACCTTACCTTAAGCCCCCTTC	6542
Qy	2161	ThrIleValThrProAsnIleValIleSerValValArgLeuAspLysGlyAsnPheAla	2180
Db	6543	ACCATCGTCAAGCCCAACATTTGTACTCTCCGATGAGCCCTTGGACAAAGGAACTTTGCT	6602
Qy	2181	GlyAlaIleLysLeuProArgTyPArgIValaLeuArgGlyGlnGlnProProAspLeuGlnThr	2200
Db	6603	GGGGCGCAAGCTGGCCCCGTGACAGAGCCCTGCTGGGAGAGAGCCCCCGGACCTTGAAGCA	6662
Qy	2201	ThrValIleLeuProGluSerValPheArgGlnThrProProValValArgProAlaGly	2220
Db	6663	ACAAGTCATTTGCTGCGAGTCTGTCTTCAAGAGAGAGCCCCCGTGTGTACGGCCGCGAGGC	6722
Qy	2221	ProGlyGluAlaGlnGlnProGlnGluLeuAlaArgArgGlnArgArgHisEProGluLeu	2240
Db	6723	CCCGGAGAGGCCCCAGAGAGCGAGAGGCTGGGACCGGACAGCGACCGGACCGGACTG	6782
Qy	2241	SerGlnGlyGluAlaValAlaSerAlaIleIleTyPArgThrIleAlaGlyLeuLeuPro	2260
Db	6783	AGCCAGGGGTAGGGCTGTGGCCAGGCTCACTCACTACCGACCTTGCGGGGGCTTACTGTGCT	6842
Qy	2261	HisAsnTyPAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr	2280
Db	6843	CATACTATGACCTTGACAGAGGCAAGGCTTGAGAGTCCCCAAAGCCCGATCATCAACACA	6902
Qy	2281	ProValIleSerIleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspLys	2300
Db	6903	CCCGGTGGAGCATCAGCGCTCATGATATATGAGAGCTTTCGCCCCCGGGCCCTGACAA	6962
Qy	2301	ProValThrValGlnPheArgLeuLeuGlnThrGlnGluArgThrGlyProIleCyVal	2320
Db	6963	CCCGTCAAGGTGCATTTCCGCTGTGGAGACAGAGAGCCGACCAACCCATCTGTCTC	7022
Qy	2321	PheThrAsnHisSerIleLeuValSerGlyTyThrGlyTyTPSerAlaArgGlyCyCysGlu	2340
Db	7023	TTCTGGAACCATTTCACTCTGTGTCAAGTGGACAGAGGTGTGCTGTGGCCAGAGGCTGTAA	7082
Qy	2341	ValValPheArgAsnGluSerHisValSerCyGlnCyAsnHisMetThrSerPheAla	2360
Db	7083	GTCGCTTCCGGAATGAGAGCCACTGTCACTCCAGTCCAAACATACATACAGCTTCGCT	7142
Qy	2361	ValLeuMetAspValSerArgArgGluAsnGlyGlnIleLeuProLeuYsThrIleuThr	2380

Dd	7143	GTGCTCATGACGTTTCTCGCGGGAGAAATGGGGAGATCTCGGCACCTGAAGACACTGACA	7202
Qy	2381	TYTATAlaleuengiValIthrlleuAlaIaleuIleuThrPhePheleuThrlleu	2400
Dd	7203	TACGTGGCTTAGGTGTCACTTGGCTGGCCCTTCTGTGTCACTTCTTCTCTCACTCTC	7262
Qy	2401	LeuAtrglIleleuAArgSerAamglnhiGlyIleArgAArgAnleuThrlAlaIaleuGly	2420
Dd	7263	TTGGATACCTGGCCCTCAACCAACAGGGATCCGAGTAACTCGAACAGCTGCCCTG3GC	7322
Qy	2421	LeuAlaIleuValIleuIleuGlyIleAamglnAlaAspleuProPheAlaCySThr	2440
Dd	7323	CTGGCTCAGCGTGGTCTTCTCTCTGGAAATCAACAGGCTACCTCTTGTGCTGACA	7382
Qy	2441	ValIleAlaIleleuLeuhiAphleuTyrlleuCySThrPheSerTrpAlaIleuIleu	2460
Dd	7383	GTCATTGCAATCCCGCTGACACTTCTGTATCCTCTGCACTTCTTCTGGGGCTCTGGAG	7442
Qy	2461	AlaIeuhIleuTyrlAArgAlaIaleuThrgluValAArgAspValAsnThrglyPrometArg	2480
Dd	7443	GCTTTGCACTGTACCGGGCACTCACTGAGGTGGCAATGTCAACCGGCCCAATGCC	7502
Qy	2481	PheTyrlThreIleuenglyTrpGlyValProAlaPheIleThrglyleuAlaValGlyleu	2500
Dd	7503	TTCTACTACATAGCTGGGCTGGGGGTGCTCGCTTCATCAAGGGCTAGCGCTGGGCTG	7562
Qy	2501	AspProgluGlyTyrlAAspProAspPheCySThrIleuSerIleTyrlAspThrlleuIle	2520
Dd	7563	GACCCCAAGGAGCTAAGGAAACCCGACTTGCTGCTGCTTCACATATAGAACAGCTATC	7622
Qy	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheIleuTyrlleu	2540
Dd	7623	TGAGTTTGTGTGGCCCGGTGGGCTTTGGCGGTCCGATGAGTGTCTTCGTACATCTG	7682
Qy	2541	AlaAlaArgAlaSerCyAlaAlaIleuArgGlnGlyPheGlnlyblybGlyProValSer	2560
Dd	7683	GCGGCCGGGGCTCTGTGTGCTGCCACAGGAGGGCTTTAGAAAGAAAGGTCTGTCTGG	7742
Qy	2561	GlyleuGlnProSerPheAlaValIleuIleuIleuSerAlaThrTrpIleuAlaIleu	2580
Dd	7743	GGCTTGACGCCCTCTTCCGCGTCTCTCTGTGTGTGAGGCCAGTGGCTGTGGCACTG	7802
Qy	2581	LeuSerValAsnSerAspThrlleuIleuPhehiSTyrlleuPheAlaThrCybAsnCyAlle	2600
Dd	7803	CTCTGTGTCAACAGCAACCCCTCTCTTCCACACTCACTCTTGTGTACCTGCAATGTGATC	7862
Qy	2601	GlnGlyProPheIlePheIleuSerTyrlValIleuSerIlybGluValArgIyValIleu	2620
Dd	7863	CAGGGCCCTTCATCTTCTCTCTTAATGTGTGCTTAGCAAGAGAGTCCGGAAGCACTC	7922
Qy	2621	LysIleuAlaCySserAArglyProSerProAspProAlaIeuthTrhIlySerThrlleu	2640
Dd	7923	AAGCTTGCTGCACCCCAAGCCAGGCCCTGACCTCTCTGACCAACCAAGTCAACCTG	7982
Qy	2641	ThrsSerTyrlAAspProSerProTyrlAlaAspIlyArgIleuTyrlGlnProTyrlGly	2660
Dd	7983	ACCTGTCTTCAACAATGCCCAAGCCCTTAGCAATAGGAGGTGGCTGTACAGCCCTTAGGA	8042
Qy	2661	AspSerAlaGlySerIleuhiAserThrsSerAArgSerIlyLysSerGlnProSerTyrlle	2680
Dd	8043	GACTCGCGCGCTCTGTGACAGCAACAGTGCCTCGGGCAAGAGTCAAGCCAGCTATC	8102
Qy	2681	ProPheIleuIleuArgIleuSerAlaIleuAsnProGlyIleuTyrlProProGlyIleuGly	2700
Dd	8103	CCCTTCTTGTGAGAGAGAGGTCCGCACTAAACCTTAGCCAAAGGCCCTCTGGCTG3GG	8162
Qy	2701	AspProGlySerIleuPheIleuGlnGlyIleuAspGlnIlehiAspProAspThrlAspSer	2720
Dd	8163	GATCAAGGACGCTGTTCTGTGAAGGTCAAGAACAGAGCATATCTGTGACAGGACTCC	8222
Qy	2721	AspSerAspIleuSerIleuAspAspGlnSerGlySerTyrlAlaSerThrlAserSer	2740
Dd	8223	GACAGTAGCCGTCTTAGAAGAGCAACAAAGTGCCCTTAATGCTCTTACCACTATCA	8282



QY 361 GIUSETYRGLNLEUTHRALVALGUALASERASPGINGLYARGASPProGlyProAArgSer 380  
 DB 1081 GATCTTACAGCTGACGGTAGAGGACAGTGAACAGGTCGGGACCGGGGTCTCGGAGT 1140  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400  
 DB 1141 ACCACAGCCGCTGTTCCTCTCTGTGTGAGGATGACAAATGATATATCCCCCAAGTTTACT 1200  
 QY 401 GIULYARGTYRValValGlnValArgGluAspValThrProGlyValAlaProValLeuArg 420  
 DB 1201 GAGAACCGCTATGTGTCCAGGTGAGGAGAGATGATCTCCAGGGCCCCAGTATCTCCA 1260  
 QY 421 ValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHisPheSerLeuMetSer 440  
 DB 1261 GTCAACAGCTCGGATGAGACAAAGGGAGCAATGCGGTGTGACTATAGCATCATGAGT 1320  
 QY 441 GJYAsnAlaArgGlyGlnPheTYRLeuAspAlaGlnThrGlyValAlaLeuAspValValSer 460  
 DB 1321 GGCATGCTCGGGGACAGTTTATCTGAGTGCACAGCTGAGCTCTGATGTGTGAGC 1380  
 QY 461 ProLeuAspTYRGLNThrThrArgGlyTYRThrLeuArgValArgAlaGlnAspGly 480  
 DB 1381 CCTCTGACTATAGAGAGACAAAGAGTACACCTTACGGGTGAGACAGATGTGTGC 1440  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspPheAsnAsp 500  
 DB 1441 CPTCCCCCACTCTATGTCTCTGCTGTGAGACAGTACAGTCTGTGATATCAAGAC 1500  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
 DB 1501 AATGCCCATCTTCGTGACAGCCCTTTCAGGCTACTGTCTGTAGAGCGCTCTTA 1560  
 QY 521 GJYTYRLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 DB 1561 GGCTACTGCTGCTCATGTCCAGGCTATCGAGCTGATGCTGTGTGACATCCCGCTCG 1620  
 QY 541 GJYTYRArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTYRGLY 560  
 DB 1621 GAATACCGCTGTGTGTGGGTGGGACATGACTTCCCTTCACATCAACATGCGCAGGC 1680  
 QY 561 TPPIleSerValAlaAlaGlnLeuAspArgGlnGluValAspPheTYRSerPheGlyVal 580  
 DB 1681 TGGATCTCTGTGGCTCTGAATCGACCGGAGAGATGATTTCTACAGCTTGGGGTA 1740  
 QY 581 GJUALAARGAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 DB 1741 GAAGCTCGAGACCATGGACTCCAGCACTCATGCTCGGCGGAGTGCAGCTGCTGC 1800  
 QY 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlnTYRThrValArgLeuAsn 620  
 DB 1801 CTGGATGTCAACGACAAACATCCAACTTTACCCCAACAGATGACACATGTGGGCTCAT 1860  
 QY 621 GIUAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 DB 1861 GAGAGTGCAGCTGTGGGACCAACGCTGTGAGCGGTCTGAGCTGTGAGCTGTCAAT 1920  
 QY 641 SerValIleThrTYRGLNLeuThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 DB 1921 AGGTGATCATCTTACAGATCAACAGTGGCAATATCTGAAACCGCTTCCATTCACAGC 1980  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTYRLeuLeuGlnArgGln 680  
 DB 1981 CAAAGGTGTGTGGGTGTGTATCTCTGTGCTGCTGCACTGACTCAAACTTAAAGGGCG 2040  
 QY 681 TYRValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 DB 2041 TATGTGTGTGTGTATCCGCTCCAGTGCATCTCGGCAAGACAGGCAAGATGTGGTG 2100  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYRThrValAsn 720  
 DB 2101 AATGTACCGAGCGCAACCACTGCTCTGTCTTCAAGAGCTCCCATATACAGTGAAT 2160

QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 DB 2161 GTTATATGAGACCGGCGCGGACGACCAACGAGTGTGTGATCAAGCGCCACCGATGAGAC 2220  
 QY 741 ThrGlyGluAsnAlaArgIleThrTYRPhMetGluAspSerIleProGlnPheArgIle 760  
 DB 2221 ACAGGTGGAATGCGCGATCACTTCAATGAGGACACACATCCCACTTCCGAGTC 2280  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTYRGLNAspGlnValSer 780  
 DB 2281 GATGAGACACGGGGGCTGTACCAACCGCTGAGCTGAGCTACGAGACCAAGTGTCT 2340  
 QY 781 TYRThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLYSerAspThrThrTYR 800  
 DB 2341 TACACCTGGCCATTAAGTCTCGGGAATGGCAATGCCCAAGTCCGACACCACTTAC 2400  
 QY 801 LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTYR 820  
 DB 2401 CTGAGATCTGTGTGAAGAGCTGATGACAAATGCCCTCAGTTCCTGCGAGACTCTTAC 2460  
 QY 821 GlnGlySerValTYRGLNAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 DB 2461 CAGGCACTGTCTATGAGAGATGTGCACCTTCACTAGCTCTGTGACATCTCAGACCT 2520  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTYRThrPheGlnGlyLYAspAspGly 860  
 DB 2521 GATCGTATCTGAGCTTAATGAGCGAGGTCTTCACTTCAAGGAGCGGACGATGGA 2580  
 QY 861 AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 DB 2581 GACGGTACTTATGTGTGATGCCAGTCAGGCAATCGGACAGCTGCGAGGCTGTGAT 2640  
 QY 881 ArgGluAsnValAlaGlnTYRValLeuArgAlaTYRAlaValAspLYSGLYMetProPro 900  
 DB 2641 CGAGAGAACGTGGCCAGTATGTCTTGGGCAATATGAGTGGACAAAGGGATCCCCCA 2700  
 QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProVal 920  
 DB 2701 GCCCGCACCTTATGAGAAATGACATCATGTGTGTGATGGAATGACAAATCCCCCTGTC 2760  
 QY 921 PheGlnGluAspGlyIlePheAspValPheValGlnGluGluAsnSerProIleGlyLeuAlaVal 940  
 DB 2761 TTTGAGCAGATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820  
 QY 941 AlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTYRGLNle 960  
 DB 2821 GCCCGGTCACAGCACTGACCCCGAGTAGAGGACCAATGCCAGATATGACAGATT 2880  
 QY 961 ValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePheSerGlyGlnLeuThrAla 980  
 DB 2881 GTGAGGGCAACATCCCTGAGGTCTTCCAGCTGACATCTTCTCCGGGAGACTGACAGCC 2940  
 QY 981 LeuValAspLeuAspTYRGLNAspArgProGlnTYRValLeuValIleGlnAlaThrSer 1000  
 DB 2941 CTGTGAGACTTGAATGAGAGACCGGCTGAGTACCTCTGTGTATCCAGGCCAGTCA 3000  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020  
 DB 3001 GCTCCTGTGTGAGACCGGCTGACATGTCAGTCCGCTCTTGAACCGGATGACACCA 3060  
 QY 1021 ProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTYRValThrAsnArgSerSer 1040  
 DB 3061 CCAAGTGTGGGCAACTTGAATCTTTCAACAACATATGACCAATCGCTCAAGAGC 3120  
 QY 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 DB 3121 TTCCCTGGGGGTGCATTGGCGAGTACCTGCCATGACCTGATATCTCAGATGATGTG 3180  
 QY 1061 ThrTYRSerPheGlnArgGlyAsnGlnLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
 DB 3181 ACTTACAGCTTGTGAGCGGGAATGAACTCAGCTGTGCTGTCAATGTCTCCACGGGT 3240  
 QY 1081 GluLeuLYleuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal 1100



Db 3241 GAGCTGAAGCTAAGCGCGCACTGGACAACAACCGGCTCTGGAGGCAATGAGCGTG 3300  
Qy 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
Db 3301 CTGGGTGTCAGAGGGGTACACAGCGTGAACCGGCTGAGTGGCGCTGGTGTGACATCATC 3360  
Qy 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
Db 3361 ACCGATGAGATCTCAACCCACAGCATCAGCGTCCGCTGAGAGACATGTACACCAAGCGC 3420  
Qy 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
Db 3421 TTCCTGTCAACCACTGAGGCTCTTCATCCAGCGGTGGCCGCAACGCTGGCCACAGCCA 3480  
Qy 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyValHisIle 1180  
Db 3481 CCGGACCAAGTGTGTCTTCAACGTACAGCGGACACAGACGCCCCCGGGGGCCACATC 3540  
Qy 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyValGlyProProPheLeu 1200  
Db 3541 CTCAACGTGAGCTGTGGTGGTGGCCAGCGCCAGCGGCCCGGGGGCGGGCCCTTCTCTG 3600  
Qy 1201 ProSerGluAspLeuGlnGluArgLeuTyrlleuAsnArgSerLeuLeuThrAlaIleSer 1220  
Db 3601 CCTCTGAGAGACTGTCAGAGAGCGCTTACACTCAACCGCAGCCTGTCAGCGCACTTCG 3660  
Qy 1221 AlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsnArg 1240  
Db 3661 GCACAGCGGTGTCTGCTCCCTTCAACAACAATCTGTGGGGAGACCCCTGCAGAACTAC 3720  
Qy 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
Db 3721 ATGGCTGTGTGTGTGGTGTGGCTTCACTCTCCGCGCCCTTTCATGCCCTCTCTCTC 3780  
Qy 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysAspProGlyPhe 1280  
Db 3781 GGTCTCTTCCGGCCCATCCACCCCGTGGAGGGCTGCGCTGCGCGCCGCGCCGCTTC 3840  
Qy 1281 ThrGluAspArgCysGluThrGluValAspLeuCysArgSerArgProCysGlyProHis 1300  
Db 3841 ACGGGGTACTACTGGAGACCGAGGGTGAACCTCTCTACTCGCGGCCCTGTGGCCCCAC 3900  
Qy 1301 GluArgCysArgSerArgGluGlyGlyTyrlThrCysLeuCysArgAspGlyTyrlThrGly 1320  
Db 3901 GGGCGCTGCCGAGCGCGAGGGCGGCTACACTTGTCTGTGTGATGCTACACGGGT 3960  
Qy 1321 GluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysLeuAsnGly 1340  
Db 3961 GAGCACTGTGAGTGTGCTGCTCAAGGCGTTCACCCGGGTGTCTGCMAAGATGGG 4020  
Qy 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
Db 4021 GGCACCTGTGTACCTGTGTGGTGGCGGTTCMAAGTGAGATGGCCATCTGGAACTTC 4080  
Qy 1361 GluArgProArgCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
Db 4081 GAGAAAGCCCTTACTGCAAGTGAACAAGCGCACTTCCCGCCCACTCTTCACTCACTTT 4140  
Qy 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGluArg 1400  
Db 4141 CGCGGCTGCGCCACAGTTCACCTTCACCTCGGCGCTCTCGTTGCCAAGAGAGCGC 4200  
Qy 1401 AspGlyLeuLeuLeuTyrlAsnGlyArgPheAsnGlyArgHisAspPheValAlaLeuGlu 1420  
Db 4201 GACGGGTGTGCTGTATACATAGGCGTTCATATGAGAACATGATTTGTGGCCCTTCAG 4260  
Qy 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
Db 4261 GTGATCCAGAGACAGTCCAGTCACTTCTCTGACGGGGAGTCAACAACAAGTGTGCC 4320  
Qy 1441 ProPheValProGlyValValSerAspGlyGlnThrPheThrValGlnLeuTyrlTyrl 1460

Db 4321 CCATTGTCGCCCGAGAGATGATGATGCCAGTGGCATATAGGTGCAAGCTGAATACTAC 4380  
Qy 1461 AsnArgProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnArgValAla 1480  
Db 4381 AATAAGCCACTGTGTGGTCAAGAGGGGTCCACAGGGCCCATCAAGACAGAAAGGTGCT 4440  
Qy 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGluSerValLeuGly 1500  
Db 4441 GTGTGACCGTGTGATGGTGTGACACAGAGATGGCTTGGCTTGTGATGTCTCTGGGC 4500  
Qy 1501 AsnArgSerCysAlaAlaGlnGlyThrGlnGlyGlySerIleGlySerLeuAspLeuThr 1520  
Db 4501 AACTACTCTGTGTGCTGCCACAGGGACCCAGGGGTGGACAGAAATGTCCTGATCTGACG 4560  
Qy 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
Db 4561 GGGCCCCCTGTATAGGGGGGTGCTGACCTGCGCCGAGACCTTCCAGTCCGAATGCGG 4620  
Qy 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
Db 4621 CAGTTCTGGGCTGCATGCGGAACCTGCAAGTGAACAGCCGACATGACATGGCTGAC 4680  
Qy 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaArgValAsnValCysAspSer 1580  
Db 4681 TTCAATGGCAACAATGGCACCGTGTCTGGCTGCCCTGCAAGAAAGATGTGTGACAGC 4740  
Qy 1581 AsnThrCysHisAsnGlyGlyTyrlThrCysValAsnGlnThrAspAlaPheSerCysGlyCys 1600  
Db 4741 AACACTTGCCACAATGGGGGACCTTGTGTGAACAAGTGGAGCGGTTCAGCTGGAGTGC 4800  
Qy 1601 ProLeuGlyPheGlyGlyTyrlSerCysAlaGlnGluMetAlaAsnProGlnHisPheLeu 1620  
Db 4801 CCTCTGGCTTTGGGGGCAAGAGCTGCGCCAGAAATGGCCATTCACAGCATCTCTCTG 4860  
Qy 1621 GlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyrlLeu 1640  
Db 4861 GGCAGACGCTGTGTGGCTGGCATGGCTCTCGCTGCCATCTCCAACTCCAGTGCCTC 4920  
Qy 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCTCATGTTCGACAGCGCGCAGCGGACAGTGTCTCTGTCAGAGGCATCAACAAGGGG 4980  
Qy 1661 ArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGlyTyrlThr 1680  
Db 4981 CGCAGACATCACTCAAGCTACAGAGGGCCACATGATGCTGAGCGTGAAGGCA 5040  
Qy 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAlaAsnAspGlyAspThrHis 1700  
Db 5041 GGGCTTCAGGGCTCTCTCTCCGCTGGAAGCAGGGCGGGCCAAATGACGGTGACTGGCAC 5100  
Qy 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
Db 5101 CATCAACAGCTGGCACTGGAGCCAGCGGGGGGCTGGCCATGTCTGTCTTGAT 5160  
Qy 1721 TyrlGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleSer 1740  
Db 5161 TATGGGACAGAGAGAGAGAGGGCACTGGGCCCGCGCTGCATGTCTGCACCTGAC 5220  
Qy 1741 AsnIleThrValGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
Db 5221 AACATTAACAGTGGCGGAATACCTGGGCGAGCGGCGGTGTGGCCGTGGCTTCGGGGC 5280  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGlyGlyValAsnSerLeuAspProSer 1780  
Db 5281 TGTTTGACGGGTGTGGGAGAGATACCCAGAGGGGGGTTAACAACCTGAGATCCAGC 5340  
Qy 1781 HisGlyLeuSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGAGCATCAACGTGAGAGCAAGGCTGTAGCTGCTGACCTTGTGATCAAC 5400  
Qy 1801 ProCysProAlaAsnSerTyrlCysSerAsnAspTrpAspSerTyrlSerCysAsp 1820  
Db 5401 CCGTGTCTGTACAGCTATGTGACAGCAAGACTGGGACAGCTATTCTGACGTGAT 5460



Qy	1821	ProGlyTYTyrGlyAspAsnCyserThrAsnValCysAspLeuAsnProCyseGluHisGln	1840
Dp	5461	CCAAGTACTATGGGACAACTGTACTATATGTGTGACTGAAACCCGTGTGAGCAACAG	5520
Qy	1841	SerValCyserThrArgLeuProSerAlaProHisGlyTYTyrThrCyseGluCysProProAsn	1860
Dp	5521	TCTGTGTACTACCCCGCAACCCCAAGTCACCCCCCATGGCTATACCTGGAGTGTCCCCAAAT	5580
Qy	1861	TytleuGlyProCotyCysGluThrArgL1alaArgInProCyserProArgGlyThrProGly	1880
Dp	5561	TACCTTGGGCACTACTGTGAGACCAAGGATGACAGCCTTGTCCCCGGGCGTGGGGGA	5640
Qy	1881	HisProThrCyseGlyProCyseAsnCyAspValSerLysGlyPheAspProAspCyseAsn	1900
Dp	5641	CATCCCACTATGGGCCCATGTGACACTGTGATGTACAGCAAAAGCTTTGACCCAGACTGCAAC	5700
Qy	1901	LysThrSerGlyGluCysHisCysLeuGluAsnHisLysThrProProGlyLysProThr	1920
Dp	5701	AAGCAACACGGCGAGTGTCCACTGTGCAGAGAAACACTACCGGCCCCAGGAGCCCCACC	5760
Qy	1921	CysLeuLeuCyAspCyAspCysTYrProThrGlySerLeuSerArgValCysAspProGluAsp	1940
Dp	5761	TGCCTCTTGTGTGACTGTCTTACCCACAGAGCTCTTGTTCAGAGTGTGTGACCTCGAGGAT	5820
Qy	1941	GlyGlnCyProCyLysProGlyValL1eGlyArgGlnCyAspArgCysAspAsnPro	1960
Dp	5821	GGCCAGTGTTCATGTCAACCAACGAGTGTATCATCGGGCGCTGATGTGACCGCTGTGACAACCT	5880
Qy	1961	PheAlaGluValThrThrAsnGlyCyseGluValAsnLysThrAspSerCysProArgAlaLe	1980
Dp	5881	TTTGTGTAGGTGACCAACAATGGCTGTAAATGAAATTAAGACGTGTGCCACAGACCAATT	5940
Qy	1981	GluAlaGlyL1eLTPTrProArgThrArgPheGlyLeuProAlaAlaL1ProCyAspPro	2000
Dp	5941	GAGGCTGGAGACTGTGGTGGCCCCGTATCCGGCTTCGGGCTGTGCTGTCTCTGTCTCC	6000
Qy	2001	LysGlySerPheGlyTYrThrAlaValArgHisCysAspGluHisValGlyTYrLeuProPro	2020
Dp	6001	AAAGGCTCTTTTGGGACGTGTGTGGCCACGTGTATGTAGCAACAGGGGTGTGCTCCCCCA	6060
Qy	2021	AsnLeuPheAsnCyserThrSerL1eThrPheSerGluLeuLysGlyPheAlaGluAsnGlu	2040
Dp	6061	AACTCTTCAACTGCACTGCATCTCACTTCTTCAAGCTGAAGAGGCTTGCTGTGAGGGCTGA	6120
Qy	2041	GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuAsnArg	2060
Dp	6121	CAGCGGATGATGACTCAGGCTCAAGTCAAGGGGCGTCCACGAGCTTACGCCCTGTCTGGCC	6180
Qy	2061	AsnAlaThrGlnHisL1eThrAlaGlyTYrPheGlySerAspValLysValAlaTYrGlnLeu	2080
Dp	6181	AAACCCACGACGACACACAGCTGGCTTACTTGGCAGCCAGCGTCMAAGGTGGCTTACACAGCTG	6240
Qy	2081	AlaThrArgLeuLeuAlaHisGlyLysThrGlnArgGlyPheGlyLeuSerAlaThrGln	2100
Dp	6241	GCCACGCGGCTGTGGGCCACGAGAGACACCCAGCGGGGCTTTGGGCTGTCTGTCCACACAG	6300
Qy	2101	AspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
Dp	6301	GACGTGCACCTTCACTGAGAAATCTGTGGGGGTGGGCAACGCGCTCTCTGACACAGCAAC	6360
Qy	2121	LysArgThrL1eTPTrGluLeuL1eGlnGlnThrGluGlyGlyTYrThrAlaTPTrLeuGlnHis	2140
Dp	6361	AAAGCGCACTGGAGCTATATCCAGAGACAGAGGGGTGCAACGGCTGTCTCTCCAGAC	6420
Qy	2141	TyrGlnAlaTYrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTYrLeuSerProPhe	2160
Dp	6421	TATAGAGGCTTACGCAAGTGCCTGTGGCCAGAAACATGGGCAACCTTACTTAAGCCCCCTTC	6480
Qy	2161	ThrL1eValThrProAsnL1eValL1eSerValValaGluLeuAspLysGlyLysAsnPheAla	2180
Dp	6481	ACCAATGTCAAGCCCAACATTTGTATCTCCGTATGTGCGCTTGTGACAAAGGAACTTTGCT	6540

QY	2181	GJyAlAlySleuProAgTyrgYlAlAleuPrGlyYglUnInProPoaSleuGLThr	2200
Dd	6541	GGGGCCAAAGCTGCCCGCCTACAGAGCCCTGGGTGGGAAGAAGCCCCCGGACTTGGACA	6600
QY	2201	ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlGly	2220
Dd	6601	AACAGCATTCCTGCCTGAGTCTGTCTTCAAGAGACGCCCGCCGTGGTCAAGCCCGCAGGC	6660
QY	2221	ProGlyGlyAlaGlnGluNPProGlnGluLeuAlaArgArGlnAlaGATGSHAProGlnUeu	2240
Dd	6661	CCGGAGAGGCCCAAGAACCAAGAGAGTGCCACAGGGAACAGCACCGCACCCGAGCTG	6720
QY	2241	SerGlnGlyGlyAlaValAlaSerValIleIleTyraGrThrLeuAlaGlyLeuLeuPro	2260
Dd	6721	AGCCACGGGTAGGGCTGTGGCCAGGCTCATCATCAACGCACCCTGGCCGGGCTACTGGCT	6780
QY	2261	HisAntyTrApProAbplyAlaArgSerLeuArGylProlyValArgProIleIeAnthr	2280
Dd	6781	CATAACTATAGACCTTAGACAAAGCCGAGCTTGAAGATCCCACAAACCCCGAATCATACACA	6840
QY	2281	ProValValSerIleSerValHisASPAspGluGlyUeuLeuProAcgAlAleuAspLys	2300
Dd	6841	CCCGTGGTAGACATCAAGCTTCATGATGATGAGAGCTTCTGGCCCGGGCCCTGGACAA	6900
QY	2301	ProValThrValGlnPheArgLeuLeuGluNthrgLynArgThrylsProIleCysVal	2320
Dd	6901	CCCGTCAAGGTCAGTTCCGCTCTCGAGACAAAGAGAGCGAACCAACCATCTGTGTC	6960
QY	2321	PheTrpAsnHisSerIleLeuValSerGlyThrdLylyTrpSerAlaArgGlyCysGlu	2340
Dd	6961	TTCGGAACCAATTAACTCTGTGTCAAGTGAACAGGTGAGTGGCCAGAGGCTGGAA	7020
QY	2341	ValValPheArgAsnGluSerHisValSerCysGlnCyASnHisMetThrSerPheAla	2360
Dd	7021	GTCCGTCTCCGCATATGAGAGCATGTCAGGTCGCAAGTGAACCATGAGAGCTTGCT	7080
QY	2361	ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuGlyThrLeuThr	2380
Dd	7081	GTGCTCATGACGTTTTCTCGCGGGGAATGGGAGATCTCGCACTGAAGACACTACA	7140
QY	2381	TyrValAlaLeuGlyValThrLeuAlaAleuLeuThrPhePhePheLeuThien	2400
Dd	7141	TACGTGGCTCAAGTGTCACTTGGCTGGCTTCTGTCTCACTTCTTCTCTCACTCTC	7200
QY	2401	LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAleuGly	2420
Dd	7201	TTGGGTATCCCTCGCCTCCAACCAACAGGGCATCCGATCCGACAGGCTGCCCTGGGC	7260
QY	2421	LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr	2440
Dd	7261	CTGGCTACGGCTGCTTCTCTCTGGAAATCAACAGGCTACCTCCCTTTGTGCTGCACA	7320
QY	2441	ValIleAlaIleLeuLeuHisPheLeuTyrcUcyThrPheSerTPRAlAleuLeuGlu	2460
Dd	7321	GTCAATTGGCANCCGCTGACACTTCTGTACTGCTGCACCTTTTCTGGGCTCTGTGAG	7380
QY	2461	AlaLeuHisLeuTyraArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Dd	7381	GCCCTTGACCTGTACCCGGGACATCACTAGAGGTGGCATGTCAACAACCGGCCCATCGC	7440
QY	2481	PheTyTyrrMetLeuGlyTPRGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu	2500
Dd	7441	TTTCACTACATGCTGGGCTGGGGCGTGCCTGTCTTATCAACAGGACTAAGCGTGGGCTG	7500
QY	2501	AspProGlnGlyTyrglyAsnProaspPheCysTrpleuSerIleTyraSPThrLeuIle	2520
Dd	7501	GACCCCAAGGGCTACGGGAACCCGACTTCTGTGCTGCTCTCCATCTATGACACGCTCATC	7560
QY	2521	TrpserPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrlleu	2540
Dd	7561	TGAGATTGTGTGGCCGGTGGCCCTTTCGGTCCANAGATGTCTTCTGTACATCTGTG	7620
QY	2541	AlaIlaArgAlaSerCysAlaAlaGlnArgGlnIdylehGlnUlyslyGlyProValSer	2560

```

Db      7621 GGGGCCCCGGGCTCTGCTGCTGCGCAGCGCGCTTGAGAGAAAGCTCTGCTGCG 7680
Qy      2561 G1yleuGlnProSerPheAlaValLeuLeuLeuSerAlaThrProLeuAlaLeu 2580
Db      7681 GGCTCGACGCTCTCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7740
Qy      2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyLeuPheAlaThrCysAsnGly 2600
Db      7741 CTCTGCTGCAAGGAGACCCCTCTCTCTCACTTCTTCTGCTGCTGCTGCTGCTGCT 7800
Qy      2601 G1nglyProPhe11ePheLeuSerTyValValLeuSerTyGluValArgLysAlaLeu 2620
Db      7801 CAGGGGCCCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7860
Qy      2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640
Db      7861 AAGCTTGCTGAGCGCGAGCGCGAGCGCGCTGAGCCCTGCTGCTGCTGCTGCTGCTG 7920
Qy      2641 ThrSerSerTyAsnCyAsProSerProTyAlaAspGlyArgLeuTyArgInProTyArgly 2660
Db      7921 ACCTGCTCTTACAGCTGCGCCAGCCCTTACGAGAGTGGCGGCTGTACAGCCCTTACGA 7980
Qy      2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyArgly 2680
Db      7981 GACTGGCGGCTCTCTGCAAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8040
Qy      2681 ProPheLeuLeuArgGlyLysSerAlaLeuAsnProGlyGlyGlyProProGlyLysGly 2700
Db      8041 CCTCTCTCTGAGGAGAGAGTCCGCTGAGAACCTTGGCGAGAGGCGCCCTTGGCGGAG 8100
Qy      2701 AspProGlySerLeuPheLeuGlyGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720
Db      8101 GATCCAGGCAAGCTGCTGTCTGAGAGTCAAGACAGAGCATGATCTTGAACAGCACTCC 8160
Qy      2721 AspSerAspLeuSerLeuGlyLysAspArgGlnSerGlySerTyAlaSerThrHisSerSer 2740
Db      8161 GACAGTGAACCTGCTCTGAGAGACAGCAGAGTGTCTCTATGCTCTTACCCACATCA 8220
Qy      2741 AspSerGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGly 2760
Db      8221 GACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8280
Qy      2761 TrpAspSerLeuLeuGlyProGlyAlaGlyLysGlyLysGlyLysGlyLysGlyLys 2780
Db      8281 TGGGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8340
Qy      2781 GlyGlyProGlyProGlyLysAlaProTyrProGlyAspPheGlyThrAlaLysGly 2800
Db      8341 GGGGGCCCGAGGCTGCGAGAGCGCCCTGCGCAGAGAGACTTGGGAGCAGCAGCAAGAG 8400
Qy      2801 SerSerGlyAsnGlyAlaProGlyLysArgLysGlyLysGlyLysGlyLysGlyLys 2820
Db      8401 AGTATGCGCAAGCGGCGCTTGAAGAGCGGCTGCGGAGATGAGAGATGCTGCTGCTG 8460
Qy      2821 G1nglySerLeuGlyProLeuProGlySerSerAlaGlnProHisLys----- 2836
Db      8461 CAGGGGCTCCCTAGGCGCCCTTCAAGGCTCTTCTGCGCAGCTCAAAAGGTGAGTGGGGC 8520
Qy      2836 ----- 2836
Db      8521 ACCCCAGCTGCGAGCTCCCTAGTACAGACCTCATACCTCATCTCTCTGTCGCG 8580
Qy      2837 -----G1yleuLysLysCysLeuProThr11e 2847
Db      8581 CACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8640
Qy      2848 SerGlyLysSerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSerArgly 2867
Db      8641 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8700
Qy      2868 SerSerAlaSerGlySerArglySerArglySerArglySerArglySerArglySer 2887

```

```

Db      8701 TCCTCCGCTAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8760
Qy      2888 LeuGlnGlyLysLeuAsnGlyValMetProLeuAlaMetSer11eLysAlaGlyThrVal 2907
Db      8761 CTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8820
Qy      2908 AspGluAspSerSerGlySerGlyLysPheLeuPhePheAsnPheLeuHis 2923
Db      8821 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8888

RESULT 5
AX367102      9321 bp      DNA      linear      PAT 16-FEB-2002
LOCUS      Sequence 21 from Patent WO0198354.
DEFINITION      AX367102
ACCESSION      AX367102
VERSION      AX367102.1 GI:18855304
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1      Griffin,J.A., Kalliock,D.A., Tribouley,C.M., Yue,H., Nguyen,D.B.,
Tang,Y.T., Lal,P., Policky,J.L., Azimzal,Y., Lu,D.A., Gaul,R.,
Yao,M.G., Burford,N., Hafalia,A.J., Baughn,M.R., Bandman,O.,
Patterson,C., Yang,J., Xu,Y., Warren,B.A., Ding,L. and
Sanjwal,M.S.
TITLE      Receptors
JOURNAL      Patent: WO 0198354-A 21 27-DEC-2001;
INCYTE      Genomics, Inc. (US)
FEATURES
source      1..9321
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /note="Incyte ID No: 6977010CB1"
BASE COUNT      1780 a 3008 c 2776 g 1757 t
ORIGIN
Alignment Scores:
Pred. No.:      0      Length:      9321
Score:      15279.00      Matches:      2893
Percent Similarity:      98.44%      Conservative:      2
Best Local Similarity:      98.37%      Mismatches:      14
Query Match:      98.29%      Indels:      32
DB:      6      Gaps:      8

US-09-916-849a-3 (1-2923) x AX367102 (1-9321)
Qy      1      MetArgSerProAlaThrGlyValProLeuProThr---ProProProProLeuLeu 19
Db      223      ATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 282
Qy      20      LeuLeuLeuLeuLeuLeuProProProProLeuLeuGlyAspGlnValGlyProCysArgSer 39
Db      283      CTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 342
Qy      40      LeuLysSerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrPheCysPro 59
Db      343      TTGGGGTCCAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402
Qy      60      SerSerAlaSerLeuLeuLeuProLeuTyThrSerArgCysArgAspAlaGlyThrGlyLeu 79
Db      403      TCCCTCAGAGTGAACCTCTGCTCTTACAGAGCGCTGAGAGATGCGGAGACTGAGACTG 462
Qy      80      ThrGlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlyLysGlyAla 99
Db      463      ACTGGCACTGTGAACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
Qy      100      His11eProLeuProProAlaProGlyGlyCysProThrSerCysArgLysLeuGlyTyr 119
Db      523      CATTTCCCTTACCAACAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582

```

QY 120 GYGLVHIELEUSERPROGLINGLYLVLEUTHLEUPROGLINGLYHISPROCYBLEUYS 139  
DB 583 GAGGACCACTTTCACCAAGGGCAAGCTCACTCCGAGAGACCCGGTCTTAAAG 642  
QY 140 AIAProAArgLEUARGCYEGINSERCYELYVLEUALGIMIALPROGLYLEUARGIAGLY 159  
DB 643 GCTCCAGGCTCAGAGTCCAGCTCTCAGAGCTGGCAAGGCCCGGGCTCAGGGCAGGG 702  
QY 160 GIUARGSERPROGLINGLYSERLEUGLYYARGARGLYVARGAANVALASNTHRALAPRO 179  
DB 703 GAAAGGTCAACGAAGAAGTCCCTGGGTGGGCGCTGGAAGAAAGAAATTAATACAGCCCC 762  
QY 180 GINPHEGLNPROBROSERTRYGINALATHRVAILPROGLUANGINPROALAGLYTHPRO 199  
DB 763 CAGTTCCAGCCCCCAGCTACCAAGGCCACAGTCCGGAAGAACCAAGCCAGGCAAGCCCT 822  
QY 200 VALAIAISERLEUARGIAIIEAAPPROAARGIUGLYVLAAGLYVARGLEUGLYTRYTHR 219  
DB 823 GTTGCAATCCCTGAGGGCCATCGACCCGAGAGGTAAGGCAAGTCACTGAGTACAC 882  
QY 220 METASPAIALEUPHEAAPPSEARASERASNGINPHEPHESELEUAPPROVALTHRGY 239  
DB 883 ATGGAGGCCCTCTTATAGCCGCTCAACCAAGTTCTTCTCCCTGAGCCCAAGTCACTGGT 942  
QY 240 AIAVALIETHRALAGIUGLYLEUAPARGIUGLYTRYLSESTHRHISVALPHEARGVAL 259  
DB 943 GCAAGTACCAAGCCCAAGAGGTGAGTCCGAGAACCAAGACCAACCTCTTCAAGGCTC 1002  
QY 260 THRALAGIAPSPHISAGLYMETPROARGARGSERALILEUALATHREUTHRIELEUVAL 279  
DB 1003 AGGCGCCAGAGACCAAGCAAGCCCAAGAGTCCCTGCTACACTCAACATCTTGCTT 1062  
QY 280 THRASPTRHASPSPHISAPPROVALPHEGLINGLINGLYTRYLYSGLUSELEUARG 299  
DB 1063 ACTGACACCAATGACATGACCTGTGTTGAGACAGAGATGACAGAGAGAGCTCAAG 1122  
QY 300 GIUASNLUGLYVALGLYTRYGLIVALLLEUTHVALARGIALATHRAPGLYAPRALPRO 319  
DB 1123 GAGAACCTGAGAGGTGCTATGAGGTGCTCACTGTCAAGGGCCACGAGTGTGATGCCCT 1182  
QY 320 PROASPAIASNIIIEUTHYRARGLEUAGLYSERGLYVSEPROSERGLYVAL 339  
DB 1183 CCGAATGCCAATATCTGTAACGCTGCTGAGAGAGATCTGAGGGCAGCCCCCTGAAAGTC 1242  
QY 340 PHEGLUIEAPPROARGSERGLYVALIIEARGTHARGIYPROVALAPARGIUGLY 359  
DB 1243 TTTGAGATGACCTCGCTCTGGGGTGAATCCGAACCCGTGGCTGTGATGGAGAAAG 1302  
QY 360 VALGLUSETRYRGLIETHRVAILGLUALASERAPGLINGLYARGAPPROGLYPROARG 379  
DB 1303 GTGGAAATCTTACCAAGCTGAGGTAAGGACAGAGTGAACAGAGTCCGGGCTCTCCG 1362  
QY 380 SETTRHTRALAIALVALPHELEUSERVALGIUASPAASPAASPAASPAASPAASPA 399  
DB 1363 AGTACCAAGCCGCTGTTTCTTCTGTAAGAGATGACATATATATGCCCCCAAGTTT 1422  
QY 400 SERGIULYARGTRYVALVALGINVALARGGIUASPVAILTHPROGLYVALAPROVALLEU 419  
DB 1423 AGTGAAGAGGCTATGTGTCCAGGTGAGGAGAGATGTGATCTCAAGGGGCCCAAGTATC 1482  
QY 420 ARGVALIETHRALASERAPARGAPPYRGYSEASPAALVALIHSYRSERTIEMET 439  
DB 1483 CGAGTCAAGCCCTCGATCGATCAAGAGGAGCAAGTCCGTGTGTGACATATAGCAATCAG 1542  
QY 440 SERGIYASPAIALARGIYGINPHETRYRLEUASPAIALGINTRGIVALALEUASPVAIL 459  
DB 1543 AGTGGCAATGCTCGGGGACAGTTTATCTGATGAGCCCAAGTCTGAGATGGG 1602  
QY 460 SERPROLEUASPTRYGLIUTHTRYLYSGIUTYTRYTHLEUARGVALARGIAGINAPGLY 479  
DB 1603 AGCCCTCTTGATCATATGAGACCAAGAGGTACACCTTACGAGGTGAGACACAGATGGT 1662  
QY 480 GLYARGPROBROLEUSERANVALSERGLYLEUVALIETHRVAILLEUASPIIEAN 499

DB 1663 GGCCGTCGCCCACTCTTAATGTCTGTGGTGGTACAGTACAGGTCTGGATATCAAC 1722  
QY 500 ASPAASPAIALPROIIEPHEVALSERTHPROPHEGINALATHRVAILLEUGLYSERVALPRO 519  
DB 1723 GACAAATGCCCATCTTGTGTGACAGCCCTTTCAGAGGTACTGTCTTGAAGAGCTCC 1782  
QY 520 LEUGIYTRYLEUVALLEUHHISVALGINALILEASPAIALAPARGIAGLYASPAASPAIAL 539  
DB 1783 TTAGGCTACCTGTGTTCTTCACTGTCAGGCTTACAGCTGATGTCTGTGACATATCCGC 1842  
QY 540 LEUGIYTRYARGLEUALAGLYVALIIEAAPPHEPROPHETRIEASPAASNGLYTHR 559  
DB 1843 CTGGAATACCCCTGCTGGGTGGGACATGATCTTCCCTTCAACATCAACATGAGCACA 1902  
QY 560 GIYTRPIIESEVALAIAIAGLYLEUASPARGLIUGLYVLAAPPHETRYSERPHEGLY 579  
DB 1903 GGTGGAATCTGTGGCTGTGAACGTGACCGGAGAGAAAGTTGATTTTACAGCTTGG 1962  
QY 580 VALGIUALIAPSPHISAGLYTHTRPROALAEUTHRALASERVALSERVALIETHR 599  
DB 1963 GTAGAACTCGAGACCAAGCAAGCTCAAGCACTGCTGCGCAAGTCAAGCTGACT 2022  
QY 600 VALLEUASPVAILASPAASPAASPROTHRPHETRGINPROGLIYTRYVALARGLEU 619  
DB 2023 GTCTCGATGTCACAGAACAAATCACTTACCCCAAGAGTACACAGTGGGCTC 2082  
QY 620 ASNGIUSAPPAIALAVALIAGLYTHSERVALIETHRVAILSERVALASPARGASPAAL 639  
DB 2083 AATGAGATGAGAGCTGTGGGACCAAGCTGTGAGCGGTGACGTGTGAGCCGTGATCT 2142  
QY 640 HISSEVALIIEETHRYRGLIIEHSESERGLYASNTHRARGASPARPHESERTIETHR 659  
DB 2143 CATAGTGCATCACTTACCAATGACAGAGTGCATATCTGAAACCCCTTCTTCAATCAAC 2202  
QY 660 SERGINSETRYGLYGLYVLEUVALSERLEUALALEUPROLEUASPTRYTRYLEUGLYARG 679  
DB 2203 AGCCAAAGTGTGTGGGCTGTATCCCTTGCCTGCACTGAGCTACAAACTTGAAGGG 2262  
QY 680 GINTYRVALILEUALAVALIETHRALASERAPGLYTHARGIINAPTRHIALGINILEVAL 699  
DB 2263 CAGTATGTGTGGCTGTATCCGCTCCAGTGGCACTCGGAGAGACCGGCAAGATGTG 2322  
QY 700 VALAENVALIETHRALASNTHRHISARGPROVALPHEGINSESETHISTRYTRYVAL 719  
DB 2323 GTGAATGCACCAAGCCCAACCAATGCTGTCTGTGAGAGTCCCAATATCAAGTGT 2382  
QY 720 AENVALAENGIUASPARGPROALAGLYTRYTHRVAILLEUIESERVALATHRAPGLU 739  
DB 2383 AATGTTAATGAGAACCGGCGGAGGACCAAGGTGTGATCAAGCGCAAGATGAG 2442  
QY 740 APPTRHGIYGIUASPAIALARGIIEHTRYRPHMETGIUASPTRIEPROGLIAPHEARG 759  
DB 2443 GACACAGTGAAGATGCCGATCACTTCACTTGAAGAGACAGATCCCCCAAGTCCGC 2502  
QY 760 IIEASPAIALAPTRHGIYVALIETHRVAILAGIUEUASPTRYRGIUASPVAILVAL 779  
DB 2503 ATCGATGACAGACAGGGGGCTGTCAACCAAGCTGAGCTGAGTCAAGAGAACCAAGTGT 2562  
QY 780 SETTRYTRHLEUALIIEETHRALARGASPAASNGIYIIEPROGLINLVSESTHAPTRHTR 799  
DB 2563 TCTTACACCTCGCCATTAAGCTCTGGGACATGAGCATTTCCCAAGAAATCCGACCAACC 2622  
QY 800 TRYLEUGIUILEUVALASPAASPVAILASPAASPAASPAASPAASPAASPAASPAAS 819  
DB 2623 TACCTGAGAGATCTGTGTGAAGAGTGAATGACATATGACATGCTCTGAGAGACTCC 2682  
QY 820 TRYGINIYSESERVALTRYGIUASPVAILPROBROPHETRSEVALLEUGINIIESERVAL 839  
DB 2683 TACCAAGGCAAGTGTATGAGAGATGTGCAACCTTCACTACAGCTCTGCAATCTCAACC 2742  
QY 840 THRASPARGASPSERGLYLEUASNGIYARGVALIETHRYTRHPEGLINGLYIYASPARSP 859

Dh 2743 ACTGATCGTGAATTCGAGCTTAATGACGGGCTCTTCAACCTTCGACGAGGCGACGANT 2802  
Qy 860 G1AaPBG1YAaPPhel1eValG1uSerThrsSerG1y1eValArgThrLeuArgLeu 879  
Dh 2803 GGAAGCGGTGACTTATTGTTGAGTCCAGTCCAGGCACTGTCGAAAGCTTACGAGGCTG 2862  
Qy 880 AaPArgG1uAaenVal1aG1nTrVal1euaArg1a1Tyra1aValAaP1yG1uMetPro 899  
Dh 2863 GATCCAGAGAAAGTGGCCCAAGTATGTCTTGGCGGCAATATGCAGTGGACAAAGGAGATGCCC 2922  
Qy 900 Pro11aArgThrProMetG1uValThrValThrVal1leuAaPValAaAaAaPProPro 919  
Dh 2923 CAGGCCCGACACCTATGAGAAATGACAGTCACTGTGTGATGTAATGACAAATCCCCCT 2982  
Qy 920 Val1PheG1uG1nAaPBG1uPheAaPVal1PheValG1uAaenSerPro11eG1yLeu1a 939  
Dh 2983 GTCCTTGAAGCAAGATGATTGATGTGTTTGTGAAAGAAAGCCCAATGGGCTTACGCC 3042  
Qy 940 Val1AaArgValThrAlaThrAaPProAaPBG1uG1yThrAaA1aG1n1leuMetTyG1n 959  
Dh 3043 GTGGCCCGGTTCACAGCACTGACCCCGATGAAAGGCAACAATGCCCAATATATGACAG 3102  
Qy 960 11eValG1uG1yAaA11ePProG1uVal1PheG1uLeuAaP11ePheSerG1yG1uLeuThr 979  
Dh 3103 ATGTGAGAGGCAACATCCCTGAGGTCTTCCAGCTGGAATCTCTCCGGGAGCTGACA 3162  
Qy 980 A1aLeuValAaPLeuAaPThrG1uAaAaPArgProG1uTyra1eVal11eG1nA1aThr 999  
Dh 3163 GCCCTGAGACTTATGACTACAGAGACCGGCTGAGTACGTCTGTATTCAGGCCACG 3222  
Qy 1000 Ser1aPProLeuVal1SerArg1aThrVal1aValAaG1eLeuAaAaPArgAaAaAa 1019  
Dh 3223 TCAGCTCCTGTGAGGCGGGCTACAGTCCACGTCGCTCTTGAACCGCAATGACAAC 3282  
Qy 1020 ProPProVal1eug1yAaAaPheG1u11eLeuPheAaAaAaAaAaThrValThrAaAaSer 1039  
Dh 3283 CACCAAGTGTGGGCAACTTGTAGATCTTTCACCAACATATGCAACATCGCTCAAGC 3342  
Qy 1040 SerPheProG1yG1yA1a11eG1yArgVal1Pro1aA1aAaPProAaP11eSerAaP 1059  
Dh 3343 AGCTTCCCTGGGGGCGCATTCGCGAGTACCTGCCATGACCTCGAATCTCAGATAGT 3402  
Qy 1060 LeuThrTySerPheG1uArgG1yAaAaG1uLeuSerLeuVal1leuAaAaAaSerThr 1079  
Dh 3403 CTGACTTACAGCTTGAAGCGGGGAATGACCTGAGCTGCTCTCAATGCTCCACG 3462  
Qy 1080 G1yG1uLeuTySerLeuSerArg1a1eLeuAaAaAaAaAaAaAaAaAaAaAaAaAa 1099  
Dh 3463 GGTAGCTGAAGCTTAAGCCGCGCACTGACCAACCGGCTCTGGAGGCAATCAATGAGC 3522  
Qy 1100 Val1eLeuVal1SerAaPBG1yVal1a1SerVal1Thr1aG1nCyA1aLeuArgValThr11e 1119  
Dh 3523 GTGCTGTGTCAAGCGGCTACAGCGCTGACCGCCAGTCCGCTGCTGTACATC 3582  
Qy 1120 11eThrAaPBG1uMetLeuThr1a1Ser11eThrLeuArgLeuG1uAaPMetSerProG1u 1139  
Dh 3583 ATCAACGATGATGCTCAACCAACAGCATCAAGCTGCGCTGAGGAGCATGACCCAG 3642  
Qy 1140 ArgPheLeuSerProLeuLeuG1yLeuPhe11eG1n1aVal1a1a1aThrLeuA1aThr 1159  
Dh 3643 CGCTTCTGTCAACCACTGAGGCTCTTCACTCAAGCGGAGGCGCCACGCTGCGCAG 3702  
Qy 1160 ProPProAaP1aVal1aVal1aPheAaAaAaG1aAaAaPThrAaP1aProG1yG1yA1a 1179  
Dh 3703 CACCGGAGCAAGTGTGTCTTCAAGCTACAGGGGACACCGAGCCCGGGGGGCGCAC 3762  
Qy 1180 11eLeuAa 1199  
Dh 3763 ATCTCAACGTAAGCTGTGTGTGGCCAGCGCCAGGGGCGGGGCGGGCGGCTTTC 3822  
Qy 1200 LeuPProSerG1uAaPLeuG1nG1uArgLeuTyra1eAaAaAaAaAaAaAaAaAaAa 1219  
Dh 3823 CTGCTCTGTAGAGACTGAGAGCGCTTATCTCAACCGGAGCTCTCTACGCGCATC 3882

Qy 1220 Ser1aG1nArgVal1eLeuProPheAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 1239  
Dh 3883 TCGGCAAGCGGCTGTGCTTCCAGCAACATCTGCTCGGAGGCTTGGAGAAC 3942  
Qy 1240 Tyra1eArgCyA1aSerVal1eAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 1259  
Dh 3943 TACATGCGCTGCTGT 4002  
Qy 1260 SerVal1eLeuPheAa 1279  
Dh 4003 TCGGTCTCTGTGCGGCTATCCACCGGCGAGAGGCTGCGCTGCGTCCCGCGG 4062  
Qy 1280 PheThrG1yAaPThrCyA1aG1uThrG1yVal1AaPLeuCyA1aAaAaAaAaAaAaAa 1299  
Dh 4063 TTCAACGGTGAATCTGCGAGACCGAAGGTGAACCTCTGCTACCTGCGGCTGTGGCCC 4122  
Qy 1300 H1eG1yArgCyAa 1319  
Dh 4123 CACGGGCGCTGCGCAGACCGGAGGGGCGGCTACACTGCTGTGTGTGTGTGTGTGTGT 4182  
Qy 1320 G1yG1uH1eSerG1uVal1SerA1aAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 1339  
Dh 4183 GGTAGGACGTGTAGGTGAGT 4242  
Qy 1340 G1yG1yThrCyA1aAaPLeuLeuVal1G1yPhe1ySerCyAaAaAaAaAaAaAaAaAa 1359  
Dh 4243 GGGGAGACCTGTGTCAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4302  
Qy 1360 PheG1uTySerProTyCyA1aG1nVal1ThrThrArgSerPhePro1aA1aSerPhe11eThr 1379  
Dh 4303 TTCCAGAAAGCCCTTACTCCAGGTACCAAGCGAGCTTCCCGGCCACTCTTCATCACC 4362  
Qy 1380 PheArgG1yLeuArgG1uArgPhe11ePheThrLeuA1aLeuSerPheA1aThrTyG1u 1399  
Dh 4363 TTTCGCGGCTGTGCGGCGAGGCTTCACTCACTGAGCTCTGTGTGTGTGTGTGTGTGTGT 4422  
Qy 1400 ArgAaPBG1yLeuLeuLeuTyra1eAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 1419  
Dh 4423 CGGAGCGGT 4482  
Qy 1420 GluVal11eG1nG1uG1nVal1G1nLeuThrPheSerA1aG1yG1uSerThrThrVal1 1439  
Dh 4483 GAGGTGATCCAGGAGGAGTGTGCAAGCTCACTCTGTGCAAGGAGTCAACCAACGCGT 4542  
Qy 1440 SerProPheVal1ProG1yG1yVal1SerAaPBG1yG1nThrPhe11eThrVal1G1nLeuTy 1459  
Dh 4543 TCCCATTCGTGCGGAGGAGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 4602  
Qy 1460 Tyra1eAa 1479  
Dh 4603 TACAAATAGCACTGT 4662  
Qy 1480 A1aVal1a1ThrVal1AaPBG1yCyAaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 1499  
Dh 4663 GCTGTGTGTGACCGGTGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4722  
Qy 1500 G1yAa 1519  
Dh 4723 GGCAACTATCTGT 4782  
Qy 1520 ThrG1yProLeuLeuLeuG1yG1yVal1ProAaPLeuProG1uSerPheProVal1ArgMet 1539  
Dh 4783 ACGGAGGCGCTGT 4842  
Qy 1540 ArgG1nPheVal1G1yCyAa 1559  
Dh 4843 CCGGAGT 4902  
Qy 1560 AaPPhel1eA1aAa 1579  
Dh 4903 GACTTCAATTCGAATGCAACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4962

QY 1580 SerAenThrCyshiBaenGlyYThrCyVaIaenGlnTrpAaPaIaBheSerCyseGlu 1599  
 Db 4963 AACCAACACTTGCACATATGGGGGCACTTGCATGAACATGGGAGCGCTTACGTGCGAG 5022  
 QY 1600 CysPProLeuGlyPheGlyGlyYssSerCyBaIaGlnGluMetAlaAsnProGlnHisPhe 1619  
 Db 5023 TCCCTCCCTGGGCTTGGGGGCAAGACTCGCCGCAAGAAATGGCCATCCACAGCACTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpYr 1639  
 Db 5083 CTGGGAGAGAGCTGTGGCTGTGGCATGGCTCTTCCCTCCATCTCCCAACCTGTGAC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 Db 5143 CTCAGCTCATATGTTCCGACAGCGCCACAGGCGGACGGGTCTGTGACGGCCATCCACAG 5202  
 QY 1660 GlyATGSerThrIleThrLeuGlnLeuArgGlnGlyValMetLeuSerValGlnGly 1679  
 Db 5203 GGGCGGAGCACTACCTTACAGCTACAGAGAGGCGACGTATGCTGAGCGTGAAGGCG 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrp 1699  
 Db 5263 ACAGGGCTTCAGGGCTCTCTCTCCGTCTGAGCGACAGGCGGGCCATGACGTGACTGG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPhe 1719  
 Db 5323 CACCATGCAAGCTGGCACTGGGAGCGCACGGGGGGCTTGGCCATCCATTCGTCTTC 5382  
 QY 1720 AspTrpGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeu 1739  
 Db 5383 GATTATGGAGAGAGAGAGAGAGGAGACCTGGGCGCCCGGTGATGCTGACCTGG 5442  
 QY 1740 SerAsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaIaArgIlePheArg 1759  
 Db 5443 AACCAACATGACATGGGGCGGAATACCTGGGCGAGCGGGGTGGCTGGCTTGGG 5502  
 QY 1760 GlyCyBleuGlnGlyValArgValSerAspThrProGlnGlyValaAsnSerLeuAspPro 1779  
 Db 5503 GGTCTGTTTCAGAGGTGTGGGTGAGCAGTACGCCAGAGGGGTGTAAACGCTGTGATCC 5562  
 QY 1780 SerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSer 1799  
 Db 5563 AACCATGGGAGAGAGATCAACGTGAGCAAGGCTGTAGCTGCTCAACCTTGTACTCA 5622  
 QY 1800 AsnProCysProAlaAsnSerYrCysSerAsnAspTrpAspSerYrSerCysSerCys 1819  
 Db 5623 AACCCGTGCTCTGCTAACAGCTATTCAGCAAGCACTGGGACAGCTATTCCTGCACTGT 5682  
 QY 1820 AspProGlyTrpTrpGlyArgAsnGlyThrAsnValCysAspLeuAsnProCysGlnHis 1839  
 Db 5683 GATCCAGGTTACTATGGTGAACACTGATCTAATGTGTGACCTGAACCCGTGTAGCAC 5742  
 QY 1840 GlnSerValCysThrArgIysProSerAlaProHisGlyYrThrCysGlnCysProPro 1859  
 Db 5743 CAGTCTGTGTGACCCGCAAGGCCAGTGGCCCCCAATGGCTATTCCTGCAAGTGTCCCCCA 5802  
 QY 1860 AsnTrpLeuGlyProTrpCysGlnThrArgIleAspGlnProCysAspProArgIleTrp 1879  
 Db 5803 AATTACCTTGGGCGCATATGCTGAGACCAAGATTGACCAAGCTTGTCCCGGTGGCTGTGG 5862  
 QY 1880 GlyHisProThrCysGlyProCysAsnGlyAspValSerIleGlyPheAspProAspCys 1899  
 Db 5863 GAGCATGCCACATGATGGCCATCACTGTGATGACCAAGAGCTTTTACCCAGACTGC 5922  
 QY 1900 AsnIleThrSerGlyGlnCysHisCysGlyGlnAsnHisGlyArgProProGlySerPro 1919  
 Db 5923 AACAAACAGAGGGGCGAGTCCACTGCAAGAGAGAACATACCGGCCCCCAGGACAGCCCC 5982  
 QY 1920 ThrCysLeuLeuCysAspCysYrProThrGlySerLeuSerArgValCysAspProGlu 1939  
 Db 5983 ACCTGCTCTGTGTGACTGCTACCCACAGGCTCTTGTCCAGATCTGTGACCTGTAG 6042  
 QY 1940 AspGlyGlnCysProCysIysProGlyValIleGlyArgGlnCysAspArgCysAspAsn 1959

Db 6043 GATGGCAGTGTTCATGCAAGCAGTGTCAATCGGGCGTCAGTGTGACCGCTGTGACAC 6102  
 QY 1960 ProPheAlaGlyValAlaThrThrAsnGlyCysGlu-----ValAsnTrpAspSer 1975  
 Db 6103 CTTTGTGTGAGGTACCAACCAATGGCTGTGAAGGGCCCTTGTGTGATTAAC----- 6156  
 QY 1976 CysProAlaAlaIleGlnAlaGlyIleTrpTrpProArgThrArg----- 1990  
 Db 6157 TGTCCCGGCCCCATGAGGTGC-----TGGCTCCAGAGAAACCTTCAGCCAGTCT 6207  
 QY 1991 PheGlyLeuProAlaAlaIleProCysProIysGlySerPhe-----Gly 2005  
 Db 6208 CAGGGGCTTCCCTGTGTGCTCCCT---GAGGCCGCGCCCTTGTGGCTTCCCTCCCGAGGG 6264  
 QY 2006 ThrAlaValArgHisCysAspGlnHisArgGlyIleTrpLeuProProAsnLeuPheAsnCys 2025  
 Db 6265 ACTGCTGTGGCCACTGTATGAGACAGGGGGGTGGCTCCCGCCAAACCTCTTCAACTGC 6324  
 QY 2026 ThrSerIleThrPheSerGlyLeuLeuYsgIyPheAlaGlyLeuGlnArgAsnGlySer 2045  
 Db 6325 ACGTCATCACTTCTCAGAACTGAAGGGCTTCGTGAGCGGCTTACAGCGGAATGAGTCA 6384  
 QY 2046 GlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHis 2065  
 Db 6385 GGCCTAGACTCAGGGCGCTCCAGAGCTAGCCCTGTCTCTGCGCAGCAGCAGCAGCAGCAGC 6444  
 QY 2066 ThrAlaGlyTrpPheGlySerAspValIysValAlaTrpGlnLeuAlaThrArgLeuLeu 2085  
 Db 6445 ACGCTGAGCTACTTCCGGAGGAGAGGTCAAGGTGGCTTACAGCTGGGCAACCGGCTGTCTG 6504  
 QY 2086 AlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThr 2105  
 Db 6505 GCCCAAGAGAGACCCAGCGGGGCTTGTGGCTGTCTGCCACACAGGACGTGACTTCACT 6564  
 QY 2106 GlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsnIysArgHisTrpGlu 2125  
 Db 6565 GAGAAATCTGCGGGGTGGGAGCGGCTCTGTGACAGAGCAACAGCGGACCTGGGAG 6624  
 QY 2126 LeuIleGlnGlnThrGlnGlyGlyThrAlaTrpLeuLeuGlnHisTrpGlnAla 2145  
 Db 6625 CTGATCCAGAGACAGAGAGGTGTGACCGCTGTGCTGCTCAGCACTAAGAGGCTTACGCC 6684  
 QY 2146 SerAlaLeuAlaGlnAsnMetArgHisThrTrpLeuSerProPheThrIleValaThrPro 2165  
 Db 6685 AGTGCCCTGGGCCAGAAATCGGGCACCTAAGCCCTTCAACATGCTCAAGCGCC 6744  
 QY 2166 AsnIleValIleSerValValArgLeuAspIysGlyAsnPheAlaGlyAlaIysLeuPro 2185  
 Db 6745 AACATTGATCTCCGATGTGGCTTGGACAAAGGAACTTGTGTGGGGCCCAAGCTGCCC 6804  
 QY 2186 ArgTrpGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThrThrValIleLeuPro 2205  
 Db 6805 CCGTTCAGAGGCGCTTCGTGGGAGAGAGGCCCGGACCTTGAGACAAACGATCTTGTGCT 6864  
 QY 2206 GlnSerValPheArgGlyThrProProValValArgProAlaGlyProGlyGlnAlaGln 2225  
 Db 6865 GAGTCTGTCTTCAAGAGACCGCCCGGTGTGAGGCCCGGAGGCGCCGAGAGGCGCAG 6924  
 QY 2226 GlnProGlnGlnLeuAlaArgArgGlnArgHisPProGlnLeuSerGlnGlyGlnAla 2245  
 Db 6925 GAGCCAGAGAGCTGGCAGCGGAGACAGGAGCGCACCCGAGAGCTGAGCCAGGAGTAAAGCT 6984  
 QY 2246 ValAlaSerValIleIleTrpArgThrLeuAlaGlyLeuLeuProHisAsnTrpAspPro 2265  
 Db 6985 GTGGCAGAGCTATCATATTCAGACCTTGGCGGGGCTTACTGCTCATATCATACCTT 7044  
 QY 2266 AspIysAspSerLeuArgValProIysArgProIleIleAsnThrProValIaSerIle 2285  
 Db 7045 GACAAAGCCAGCTTGAAGTCCCGCAAGCCCGGATCAACACACCGGTGTGTGACATC 7104  
 QY 2286 SerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspIysProValThrValGln 2305

Db 7105 AGCTTCATGATGATGAGGAGCTTCTGCCCCGGGGCCCTGAGCAAAACCCGTGACGGGTGAC 7164  
 Qy 2306 PheATgLeuLeuGluuThrGluGluArgThyLysProIleCysValPheThrPheHisSer 2325  
 Db 7165 TTCGGCTGCTGGAGACAGAGAGCGGACCAAGCCATCTGTCTTCTGGAAACCATTTCA 7224  
 Qy 2336 IleLeuValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgPhe 2345  
 Db 7225 ATCTGGTCAAGTGGACAGAGTGTGTGTGGCCAGAGGCTGTGAAGTCTTCTCCGCAAT 7284  
 Qy 2346 GluSerHleValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspVal 2365  
 Db 7285 GAGAGCCAGTCAAGCTGACAGTGCACCAATGACAGCTTGTGTGTCTCATGAGACGTT 7344  
 Qy 2366 SerATgATgGlu-----AsnGlyGluIleLeu 2374  
 Db 7345 TCTCGCGGGAGTGGGGGCCACAGGGGAGCGTGCAGAGCCGTGAAATGGGAAATCTCTG 7404  
 Qy 2375 ProLeuLysThrLeuThyValAlaLeuGlyValThrLeuAlaAlaLeuLeuThr 2394  
 Db 7405 CCACCTGAAGACCTGACATACCTGGCTCTAGGTGTCACTTGGCTGCTTCTGCTCAAC 7464  
 Qy 2395 PhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyTleArgArgPhe 2414  
 Db 7465 TTCTTCTTCTCACTCTGTGGTATCTGTGCCTCCACCAACAGCGCATCTCGACGTAAAC 7524  
 Qy 2415 LeuThraAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuGlyTleAsnGlnAlaAsp 2434  
 Db 7525 CTGACAGCTGGCTGGGGCTGGCTCAGCTGGTCTTCTCTGGAAATCAACAGGCTGAC 7584  
 Qy 2435 LeuProPheAlaCysThyValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPhe 2454  
 Db 7585 CTCCCTTGTGCTGACAGTCACTGCACTCCCTGCGACCTTCTTACCTCTGCACTTT 7644  
 Qy 2455 SerTPAlaLeuLeuGluAlaLeuHisIleuThyArgAlaLeuThrGluValArgAspVal 2474  
 Db 7645 TCTTGAGCTTGTGAGAGCTTGGACCTGTACCGGGACCTGACAGAGTGGCCATATCTC 7704  
 Qy 2475 AsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr 2494  
 Db 7705 AACACCGGCCCCCATCGCTTCTACTACATGCTGGGGGTGGCTGGCTTCAATCA 7764  
 Qy 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSer 2514  
 Db 7765 GGGCTAGCGCGTGGGCTGAGACCCCGAGGGCTACGGGAAACCTGACTTCTGCTGCTCC 7824  
 Qy 2515 IleTyrAspThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSer 2534  
 Db 7825 ATCTATGACAGCTCATCTGAGATTGTGTGGCCCGGTGGCTTGTCCGTCTCGATGAGT 7884  
 Qy 2535 ValPheLeuTyrIleLeuAlaIlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlu 2554  
 Db 7885 GTCTTCTGTATCATCTCGCGGCGCGGGCTCTGTGTGGCCCAAGCGGCAAGGCTTTAG 7944  
 Qy 2555 LysLysGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAla 2574  
 Db 7945 AAGAAAGGTCTGTCTGGGCTGACGCCCTTCTGCGCTCTCTCTGCTGAGCGCC 8004  
 Qy 2575 ThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPhe 2594  
 Db 8005 ACGTGGCTGCTGGACCTGCTCTCTTCAACAGCAACCTTCTTCACTCACTCTT 8064  
 Qy 2595 AlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLys 2614  
 Db 8065 GCTACTCTCAATTGATCCAGGGCCCTTCACTTCTCTCTATGTGTGTCTTAGCAAG 8124  
 Qy 2615 GluValArgLysAlaLeuLysLeuAlaCysSerAspGlyProSerProAspProAlaLeu 2634  
 Db 8125 GAGGCTCCGAAAGCACTCAAGCTTCTGACAGCCCAAGCCCGACCTGCTCTG 8184  
 Qy 2635 ThrThyLysSerThrLeuThySerSerTyrAsnCysProSerProTyrAlaAspGlyArg 2654  
 Db 8185 ACGACCAAGTCCACCTGACCTGTCTTACAACTGCCCAAGCCCTTAGCGAGATGGCGG 8244

Qy 2655 LeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLys 2674  
 Db 8245 CTGTACAGACCTTACAGGAGACTCGGCCGCTCTTGCACAGACCAAGTGGCTGGGCAAG 8304  
 Qy 2675 SerGlnProSerTyrIleProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGln 2694  
 Db 8305 AGTCAGCCAGCTACATCCCTTCTTGTGTGAGGGAGAGTCCGCACTAAACCTTGGCCA 8364  
 Qy 2695 GlyProProGlyLeuGlyAspProGlySerLeuPheLeuGluGluGlnAspGlnHis 2714  
 Db 8365 GGGCCCCCTGGCTGGGGGATCCAGGACCTGTCTTCTGGAAGTCAAGACAGACGAT 8424  
 Qy 2715 AspProAspThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerGlySerTyr 2734  
 Db 8425 GATCTTGAACAGCACTCCGACAGTGACTTCTTGAAGACGACCAAGTGTGCTCTTAT 8484  
 Qy 2735 AlaSerThrHisSerSerAspSerGlyGluGluGluGluGluGluGluGluAla 2754  
 Db 8485 GCTCTACCCACTCATCAGACAGTGAAGAGGAAGAGAGAGAGAGAGAGAGAGCGCC 8544  
 Qy 2755 PheProGlyGlyGluGlyTyrPheAspSerLeuLeuGlyProGlyValArgLeuProLeu 2774  
 Db 8545 TTCCCTGAGAGCAGGCTGGAGTACCTGTCTGGGGCTTGAAGACAGAGACTGCCCTG 8604  
 Qy 2775 HisSerThrProLysAspGlyGlyProGlyProGlyLysAlaProTyrProGlyAspPhe 2794  
 Db 8605 CACAGTACTCCCAAGATGGGGGCCCAAGGCTTGGCAAGGCCCTTGGCCAGAGACTTT 8664  
 Qy 2795 GlyThrThrAlaLysGlySerSerGlyAsnGlyValaProGluGluArgLeuArgGluAsn 2814  
 Db 8665 GGGACCAAGCAAAAGAGAGTGAAGGCAACGGGGCCCTGAGAGAGCGGCTCGGGAGAT 8724  
 Qy 2815 GlyAspAlaLeuSerArgGlyGlySerLeuGlyProLeuProGlyLysSerSerAlaGlnPro 2834  
 Db 8725 GGAATGCTCTGTTCAGAGGGGTCTTGGGCCCTTCCAGGCTCTTGTGCCAGGCT 8784  
 Qy 2835 HisLysGlyIleLeuLysLysCysLeuProThrIleSerGlyLysSerSerLeuLeu 2854  
 Db 8785 CACAAAGGCATCTTAAAGAGAGACTGTCTGCCACATCACAGAGAGAGAGAGAGCTCTG 8844  
 Qy 2855 ArgLeuProLeuGluGluGlySerGlySerArgLysSerSerAlaSerGlyLysSer 2874  
 Db 8845 CGGCTCCCCCTGGAGCAATGACAGGGGTCTTCCGGGGCTCTCCGCTAGTGAAGGGCAGC 8904  
 Qy 2875 ArgGlyGlyProProProAspProProProArgGlnSerLeuGlnGluLeuAsnGly 2894  
 Db 8905 CGGGGCGGCCCTTCCCCCGCCACCGCCCGGCGAGAGCTTCCAGAGAGAGCTGAACGGG 8964  
 Qy 2895 ValMetProIleAlaMetSerIleLysAlaGlyThrValAspGluAspSerSerGlySer 2914  
 Db 8965 GTCATGGCCATCCGCAAGAGATCAAGGAGGACCGGTGATGAGACTGCTGAGGCTCC 9024  
 Qy 2915 Glu 2915  
 Db 9025 GAA 9027

RESULT 6  
 AB028499 8950 bp mRNA linear ROD 06-SEP-1999  
 LOCUS Mus musculus mRNA for Flamingo 1, complete cde.  
 DEFINITION AB028499  
 VERSION AB028499.1 GI:5832710  
 KEYWORDS Flamingo 1.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (slices)  
 Ueul, T., Shima, Y., Shimada, Y., Hirano, S., Burgess, R. W.,  
 Schwartz, T. L., Takeichi, M. and Uemura, T.  
 Flamingo, a seven-pass transmembrane cadherin, regulates planar  
 cell polarity under the control of Fz12



JOURNAL Cell 98 (5), 585-595 (1999)  
 MEDLINE 99418630  
 PUBMED 10490098  
 REFERENCES 2 (bases 1 to 8950)  
 AUTHORS Uemura, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUN-1999) Tadaehi Uemura, Kyoto University, Faculty of Science, Department of Biophysics, Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 606-01, Japan  
 (E-mail: t.ueamura@take.biophys.kyoto-u.ac.jp, Tel: 075-753-4195, Fax: 075-753-4197)

FEATURES  
 Source Location/Qualifiers

1. 8950  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1. 8950  
 /gene="flamingo 1"  
 1. 8763  
 /gene="flamingo 1"  
 /codon\_start=1  
 /product="flamingo 1"  
 /protein\_id="BAA84070.1"  
 /db\_xref="GI:5832711"

/translation="MRTASAPLPPLPLLLLLPPSLGDPQVPCSLSGG  
 /SSSGACAPVGMCLCPASNLMDYTRCRESGIELTGHVPHHDGLRWCPSGHAIP  
 LPPSSRGCPWSCRLIGIHLSPQTLTPEHPCLAPLRCSCKLAQAPLRAGE  
 GSPESLIGRRKKNVNTAPQOPSYOATVPENOPAGTSASLRALDIPDEGAGLEY  
 TMDLPDSRNHPFLSDPLTGVVTTAELDRFETKTHVRVRAODHGMRRALATLT  
 ILVTDNRDHPVEQOEYKESLAEENLEVEYEVLTTRATDGDAPFNANILYRLLEGAD  
 SPDAEIDPRSGVIRTRGPVDRVEVSTLTVEASDQGDGPPRSSTALVPLSVEDD  
 NDAPQSEKRYVQVAREDDTGAIVLAVASDNGSNALVYSIMSGARQOFLD  
 AQTGALDVVSPLVETTKETTLRAODGGRPLSNVSGILVYQDINDIPILVS  
 TPOQATLBSVPLVAVQALDADAGNARLESLAGAGHPPTLNGNGMISVA  
 AELDRREVDVSGVARDHCPALTRASASVTLLDNDNPTPPOPYRYLND  
 AVGTSVYVSAVDRBAHSVITTYITISGTRNRPSTISQSGELVSLPLDYLEKD  
 VLAIVASDGRQDTQIVANVTANTHREPVQSHVTVNENRERAPVTLISATDE  
 DTENARITYEMEDSIPQRI DDDTGAVTYQAEIDYEDVSYTLATADNGLPQSD  
 TYLEILVNDVNDNAPQPLRDSYQGVYEDVPPETSVLQILADPSGLNGRVTYQ  
 GGDGDGDFIVESTSGIVTRLRLDRENVAVILRAVADKMPPARTMEVTVLD  
 GNNPVPVEDEDFVENSPIGLAVARTATDPDGNATQIMQIYIGNIPVETLD  
 DIFSGRLTLVDLDYEDRPREYLVIOATSLAVSRATVYRLDRNDPNTVGNPIL  
 FNNYVNRSSPPGAGIGRYPADPDISLTYSFPERNELSVTLNMASTGERLSRA  
 LDNNRPLEAMSVLSDGVHSAVQCSLRATITIDMLTHSITLRLMDSPEKFLSPL  
 LGLIQAVALTATLPDHHVFNVDPTAPAGHILNVLSVQPPGPGGEPPLPSE  
 DLOERLYNLSLTALSAKAVLPFRQHLRBCENYMCVSLARFDSAPFLASISV  
 LFPRIHVGGLRCRCPGLTGDYCEVEVDLCYRGTGPGHRCRSGGTCLCRGCT  
 GEGCASTSGRCIPGVCNKGTCVNLVNGIKCDPCSGHREKPGQVTRSPAPRP  
 ITRGHORHPHTLASPRATKERNGLLNGRPNRKHDPALAEVIOEQVUTPSAGS  
 ITTVSPVPGVSDGQMTVQLKTYNPELDQTLQGSBQVAVSVSDGCTGVAL  
 RFQAMLGANSCAQGTQSGSKSLDLGPELDPSFVNRKAVSVSDGCTGVAL  
 DSRHIDMADFIANNQVPGCPYKICVDSISCHNGGTQVQNTPTCEBPLGGGKSC  
 AOSMANPDRFGLSSLVAMHGLVPIISOPHILNMPRTROADGTLCAVTRGRTTLQ  
 LRAGVRLSMBEGTGLASLTLRGRALQMDHMAQLGASGRHILNLSRYQGT  
 ABGRULPRHLGHLNITVCGVPGPAGVARGRCLOQVRYETPBGHSLDPSKGE  
 SLNVEFGCSLPDPCDSNPCTNSYSDNWSYSCCLVGYGNCVNDLNECHOS  
 VCTRKNTPHGYICECLPNLAPYCEPRLDQPCRGWMPGTGPCNCAVSGKQDFC  
 NKTSGCHCKEKYRPSPGPTCLCDQCYPTGSLRYCDPDDCCPKQVPCGCDRC  
 DNPFAYVTNGCENYDSCPRALBAGIMRTRFGLPAAPACKSGFVARGHCDHR  
 GMLPMLFNCTSVTSPSELKGAERLQNSGLDSGRSOLALLNNAOHTSGYPSD  
 VKYAYOLATLLAHESQRFGLSATODHPTENLARSALDLDAKRMELIOQTS  
 GATMLQHTBATASALQNMRTYLSPTITVNTVIVSVRLDKNPGFGLTSEVA  
 LRGERPDVETVILPESPREMSPWRSAGPQBTREIVLGRORRHPLESGEVA  
 SVIYHTLAGLPHNDPDRSLRPRPINTPAVISVYHDEBELPALDPLPVQ  
 FRLEETERTKPICVFNHSLIVSGTGMARGCEVVFRRSBSQCCNMHSFALM  
 DMSRRENGEILPLKTLTYVALVGLTALAMTLPLFLTLRLSRNGIRNRLAALG  
 AOLVFLIGNOALDPACTVYIILHRTLYCTPSWALRHLXRLTYRVDNVAEM  
 RFTYMLGCVPAPIITGLAVGLDPRGTPNPPCMLSTYDTLINSFAPVAPAYSMRFL  
 YILSARASCAHQGFQEKKPVSGLSLFTVLILSATWMLILVNSDTLPHYIFA  
 ACNCVQGPFIPLSVYVLSKEVRALKAKACRSPDPALTKTTLTSLNSPBPYAD  
 RLVPYQDSAGLSHSASRSKQSPSYIPLLRBSSTLNPAPVPGGLADPSGLLEQQA  
 QOHPDSDSDSLIEDDSGASVASTSSTSSSBEELNAPGCGGCTGAPRSLP  
 LHS7PKDGGSGKVPWLGDFCTTTKENSAGALBERPRNGALTRBSGLPLPSPS

TOPHKLKXKCLPTISSEKSLRLPLEQGTSSSGSSISBGRHPPPPRPSLQ  
 EQLNGVMVAMNSINAGTVDSDSGSRFLFFNFH"  
 BASE COUNT 1793 a 2813 c 2512 g 1830 t 2 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 8950  
 Score: 14647.50 Matches: 2761  
 Percent Similarity: 96.24% Conservative: 54  
 Best Local Similarity: 94.39% Mismatches: 104  
 Query Match: 94.23% Indels: 7  
 DB: 10 Gaps: 2

US-09-916-849a-3 (1-2923) x AB028499 (1-8950)

QY 1 MetATGserProAlaThnGlyValProleuProThrProProProProleuLeu 20  
 Db 1 ATGCGGACCGGGGCTGCCAGCGCCCTCCCAAGCGCTGCTGCGCTTACTGCTG 60  
 QY 21 LeuLeuLeuLeuProProProleuGlyAaPGLNValGlyProCyAaRserLeu 40  
 Db 61 TTGCTGCTGCTGCCAGCGCTGCCACTTACTGGGAGATCAAGTGGGCGCTGCTGCTG 120  
 QY 41 GlySerAaGlyAaGlySerSerGlyAaCyAaAaProMetGlyTTPLeuCyAaProser 60  
 Db 121 GGATCCGGGGGAGCGAGCTCTCTGGGGCTGCGCGCCCGTGGCTGCTGCCAGCC 180  
 QY 61 SerAlaSerAaLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeuTTPLeu 80  
 Db 181 TTGCTGCTGACCTTGGCTTACACAGCGCTGCAAGAAATCGGCACTTGAAGCTGACC 240  
 QY 81 GlyH1sLeuValProH1sAaPGLNValTTPCyAaProGluSerGluAaH1s 100  
 Db 241 GGCACCTGCTGCCACAGATGGCTGAGGGTTGGTGTCCAGAAATCAAGGGCTCAT 300  
 QY 101 IleProleuProProAlaProGluGlyCyAaProTTPserCyAaLeuLeuGlyIleGly 120  
 Db 301 ATCCCTCTTCCGCAATCTCTGAAAGCTGCGCCCTGAGCTGCTGCTGCTGCTGCTG 360  
 QY 121 GlyH1sLeuSerProGluGlyLeuLeuThrLeuProGluGluH1sProCyAaLeuYsa1a 140  
 Db 361 GGAACCTTTTCCCAAGAGCAGCGCTGACCTGCAAGAGACACCTTGTAAAGGCC 420  
 QY 141 ProATGLeuAaCyAaGluSerCyAaLeuAaGluAaGluAaProGlyLeuAaGluAaGlu 160  
 Db 421 CCAACGGCTCAAGATGCAAGCTTGGCAAGCTGGGCAAGGCCCTCAGGGCTCAGGGGAA 480  
 QY 161 ArgSerProGluGluSerLeuGlyValAaAaGlyAaAaAaValAaThrAlaProGlu 180  
 Db 481 GGCTACCAAGAAATCTCTGGTGGGCGCAAGAAAGAAATGTAATACAGCTCCCAAG 540  
 QY 181 PheGluProProSerTyrGluAaThrValProGluAaGluAaGluAaGluAaGluAaGlu 200  
 Db 541 TTCAGAGCTCCCAAGCTTACAGGCTCAGGCTGCAAGAGCTCAGGCTGCTGCTGCTG 600  
 QY 201 AlaSerLeuAaGluAaIleAaProAaPGLNValGlyAaGlyAaGlyAaGlyAaGlyAaGly 220  
 Db 601 GCATCCTTAAAGCAATGATCAATGAGGCTGAGGGGCTGAGCTTGAAGTACCAATG 660  
 QY 221 AaPAlaLeuPheAaSerAaRserAaGluAaPheSerLeuAaProValThnGlyAa 240  
 Db 661 GAGGCCCTTTGATAGCGCGCTCAATCATTTCTCTTGAACCAATCAACCGGCTGT 720  
 QY 241 ValIThrThrAlaGluGluLeuAaPGLNValThnGlyAaSerThrH1sValPheAaGlyAa 260  
 Db 721 GTCACACAGCTGAGAGCTGATCGGAGAGCAAGAGACCAACGCTTCAAGGGTCACT 780  
 QY 261 AlaGluAaPheAaGlyAaPheProAaRserAaAaGluAaThrLeuThrIleLeuValThr 280  
 Db 781 GCAACGAGTACAGGATGCGGAGAGTGGCTTGGCCACGCTTCAATCTTGGTGAACC 840  
 QY 281 AaPThrAaAaH1sAaAaProValPheGluGluGluGluTyrTyrGlyGluSerLeuAaGlu 300



Db 841 GACACCAAGATCAGACCCCTTTTGAACGACCAAGTACAGAGACCTCAGGGAG 900  
 Qy AsnLeuGluValGlyTyrGluValLeuThrValAlaGlyAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGAGAGGTGGCTATGAGGTGCTTACAGTCAAGGCGCACCGAGCGGCCCTTCCC 960  
 Qy AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGlyValAlaPhe 340  
 Db 961 AATGCAACATTCGTGTACCGCTGCTGAGGGGGGAGACAGCCCTCCAGACCCCTTT 1020  
 Qy GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
 Db 1021 GAGATGATCTCTCGTCCGGGGGTGATCCGACCCCGGCCCTGTAGACCGGGAGAGATG 1080  
 Qy GluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 1081 GAATCTTCAAGATTACAGTGGAGCGAGTACCAAGGGCCGGGACCCAGGCCCGGAGT 1140  
 Qy ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
 Db 1141 TCCACAGCCATGTGTTCTCTGCTGAGAGATGATTAATGACCAAGCCCCCAAGTTTAC 1200  
 Qy GluIleAspArgTyrValValGluValArgGluAspValThrProGlyValaProValLeuArg 420  
 Db 1201 GAGAAAGCTTATGTGTCTCAGGTGCGGAGAGATGACCCCAAGAGCCCAAGTTCTCGA 1260  
 Qy ValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaIleTyrSerIleMetSer 440  
 Db 1261 GTCAACCGCTCCGANTAGAGACAGGGGACGATGCCCTTAGTGACCTACAGATCAAGAT 1320  
 Qy GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 Db 1321 GGCATATGTCGGGGGCAAGTTCATCTGATGCTCAGACCTGAGCCCTGAGTGTGTAAGC 1380  
 Qy ProLeuAspTyrGluThrThrLeuGluTyrThrLeuArgValArgAlaGluAspGly 480  
 Db 1381 CCATCTGACCTATGAGACCAACAAAGATTAACGCTTCGAGATCCGGGACCAAGCGTGGC 1440  
 Qy ArgProProLeuSerAsnValSerGlyLeuValThrValGluValLeuAspIleAsnAsp 500  
 Db 1441 CGTCCCTCACTCTCAATGTCTCTGGCTTAGTAACAGTGCAAGTCTTAGATCAAGAT 1500  
 Qy 501 --- AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
 Db 1501 ATACGGCCCCCATCTTGTGACGACCCCTTCCAGGCCACCGTCTAGAGAGTGTCCCT 1560  
 Qy LeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 Db 1561 TTAGGCTACCTTGTCTTCAAGTCCAGGCAATTGATGCTGACGCTGGATTAACGCCCGC 1620  
 Qy LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyr 559  
 Db 1621 CTCGAGTATGACCTGGGAGGGGTTGGGCAACCTTCCCTTCAACATTAAACAATGGACA 1680  
 Qy GlyTyrPheSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 Db 1681 GGCAGGATCTGTGGCCGACAGATGGATGGGAAGGTTGATTCTACAGCTTGGGT 1740  
 Qy ValGluAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 Db 1741 GTAGAGGCCCGGAGCAATGACACCCCAAGCTTCAAGGCTTCAAGAGTACAGGTAAACC 1800  
 Qy ValLeuAspValAspAspAspAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
 Db 1801 ATCTCGAGCGTCAATGACCAACCACTTCAAGCGACGAGTACACCGTTCCGGCTC 1860  
 Qy AsnGluAspAlaAlaValGlyTyrSerValValThrValSerAlaValaAspArgAspAla 639  
 Db 1861 AACAGAGAGCGCGCGGGGACCAAGTGTGTGACCGTCTGACCGGTGAGTACGACGCT 1920  
 Qy HisSerValIleThrTyrGluIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 Db 1921 CACAGTGTCAATTACTTACCAATCAATCAAGCGGCAACCCGCAACCGCTTCTATACAC 1980

Qy SerGluSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluAlaArg 679  
 Db 1981 AGCCAAGTGTGTGATCTGTCTTCCCTTACCGCTTAGCTAGACTACAACTGAGCGG 2040  
 Qy GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleVal 699  
 Db 2041 CAGTATGTGTGGCAGTACGCTCCTCAGATGGCACAGGCGAGACAGCTCAAGTATGTA 2100  
 Qy ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
 Db 2101 GTGATGTCAACCGATGCCAACACCCAGCTCCGCTTCCAGACCTCCCACTACACAGTTC 2160  
 Qy AsnValAsnGluAspArgProAlaGlyTyrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2161 AACGTTATGAAAGCCCGGACAGGACCAACAGTGTCTGATCAAGCCCAAGATGAG 2220  
 Qy AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2221 GACACGGGGGAGATGCCGATACCTACTTATGAGATAGATCCCCCAAGTTCGA 2280  
 Qy IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 Db 2281 ATCGACGGGACACTGGGGCTGTCCACCAAGGCTGAGCTGACAGAGACAGGTG 2340  
 Qy SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
 Db 2341 TCTTATACCTCTGGGCATCACTGCTCGGACAAATGGCATTTCCCAAGATCTGACACACC 2400  
 Qy TyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 2401 TATTGAGATCTCTGGTGAATGACGTGAACGACACCCCGGCTTCTCAGAGATTCC 2460  
 Qy TyrGlnIleSerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 2461 TACAGGGACCGCTTACGAGAGAGTGCCTTCCACAGCGCTGACAGATCTTAGCC 2520  
 Qy ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnIleGlyAspAsp 859  
 Db 2521 ACGGATGAGACCTCGGCTCGAATGGCAGGGTTTCTACACCTTTCAGAGAGAGATGAT 2580  
 Qy GlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLeu 879  
 Db 2581 GGGGACGGTCACTTATTTGATGAGTGCACATCGGGCATTTGTGGCACATCGCGGAGCTG 2640  
 Qy AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValaAspGlyMetPro 899  
 Db 2641 GATCTGAGAACGTGGCCAGTACGTTCTGCGGACCTATCGGGTGAACAGGCGATGCCA 2700  
 Qy ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
 Db 2701 CAGGCCGACACCAACATGAGAGTACAGTACTGTCTGATGGAGATGACAAACCCCTC 2760  
 Qy ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 2761 GTCTTGAACAGATGATGTTGACGTATTGTGGAAGAGAAGCCCATCGGGCTGGCC 2820  
 Qy ValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 2821 GTGGCCGGGATCACAGCACTGACCAAGATGAAGAGCAACGACAGATCAATGATACAG 2880  
 Qy IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 2881 ATTGTGAGGGCAATATCCCTGAGGCTTTCAATGGAATATATCTCTGGCGAGCTACT 2940  
 Qy AlaLeuValaAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 2941 GCCCTGTATATTGACATATGAGACCGGCTGAATATATGCTCTGTCAACCGGCTACG 3000  
 Qy SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3001 TCTGTCTCCTTGTGTAGAGGGCTACTGTCCATGTCTCCGCTCTTACCGCATGATAC 3060

QY 1020 ProProValLeuGIYAsnPhelGluIleuPheAsnAsnTYrValThrAsnArgSerSer 1039  
 Db CGCCCGGCTGGGCAACTTGAAGATCTCTTAAACAATTATGACCAACCGCTCCACC 3120  
 QY 1040 SerPheProGlyGIYAlaIleGIYArgValProAlaHisAspProAspIleSerAspSer 1059  
 Db ACCTTCCCGGAGTGAGCCATAGGCGGTGTGCTCCCAAGACCCCGATATCTCAAGACGC 3180  
 QY 1060 LeuThrTYrSerPheGluIuArgGIYAsnGIYLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 Db CTGACGTATACGCTTCAAGGAGGAAATGAATCACTGCTGTCTCTCAATGCTCCACC 3240  
 QY 1080 GIYGIYLeuLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGIYAlaIleMetSer 1099  
 Db GGTGAGCTGAGACTGAGCCGGGCACTGACCAACACCGGCCCTTGGAGCCATCATGACT 3300  
 QY 1100 ValLeuValSerAspGIYValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db GTGCTGTGTCAAGATGTGTCCACAGTGTGACAGCCCAAGTGTCTGCTACAGCTCACATC 3360  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrIleuArgLeuGIYAsnMetSerProGlu 1139  
 Db ATCAAGAGAGATGTCTACACACAGCATACGCTGCTGTGAAAGACATGTCTCCAGAA 3420  
 QY 1140 ArgPheLeuSerProLeuLeuGIYLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
 Db CGCTTCTCTCAACACTGTGGGACTCTTCACTTACAGCCGTGGGCAACCTGTGGCCACA 3480  
 QY 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGIYGIYHis 1179  
 Db CCCCCGAGCATGTGTGTCTTCAATGTGCAAGAGATATGATGCCCCAGGCGGCAT 3540  
 QY 1180 IleLeuAsnValSerLeuSerValGIYGIYLeuProProGIYProGIYGIYProProPhe 1199  
 Db ATCTTCAAGCTAGAGCTGTGTGTGTGGCCAGCCCTCCAGAACCCGGGGTGGGCCACCTTT 3600  
 QY 1200 LeuProSerGIYAspLeuGIYArgLeuTYrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 Db CTGCTTCAAGAGATCTCCAGAGAGCGCTGTACTCAACCGAGCGTGTCTCAACCGGCATC 3660  
 QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAs 1239  
 Db TCGGCGAAGCGGTGTCTCCCTTCCAGACAAATTTGCTG-CGGAGACCTGTGGAGAA 3719  
 QY 1239 nTYrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSe 1259  
 Db 3720 TTACATGCAATGTGTCTGTGTGTGGGTTCACATCTCTTGGCCCTTATGTGCTTCTC 3779  
 QY 1259 rSerValLeuPheArgProIleHisProValGIYGIYLeuArgCysArgCysProProGI 1279  
 Db 3780 TTCGGGTCTTCCGGCCCATCACTCGTCGAGAGGTCTGGCTGTGCTGCCCAACCCGG 3839  
 QY 1279 yPheThrGIYAspTYrCysGIYThrGIYValAlaAspLeuCysTYrSerArgProCysGIYPr 1299  
 Db 3840 CTTAACAGCGACTACTGTGAGACCGAGGTGAGACTCTGTACTCAAGAACCTGTGGACC 3899  
 QY 1299 OHIsGIYArgCysArgSerArgGIYGIYGIYTYrTYrCysLeuCysArgAspGIYTYrTh 1319  
 Db 3900 CCATGGCGCTGTGGGAGCGAGGAGGTGTGTACTCTGTGTGTGTGTGTGTGTGTGTGT 3959  
 QY 1319 rGIYGIYHisCysGlnValSerAlaArgSerGIYArgCysThrProGIYValCysAlaHis 1339  
 Db 3960 GGGCGAGCACTGTGAAGCAAGTACCACTCAGGCCCTGTGTACTCCAGGTGTGTGTGTGT 4019  
 QY 1339 nGIYGIYThrCysValAlaAsnLeuValGIYGIYPhelCysArgCysProSerGIYAs 1359  
 Db 4020 CGGGGTACTGT 4079  
 QY 1359 pPheGIYArgProTYrCysGlnValThrThrArgSerPheProAlaHisSerPheIleTh 1379  
 Db 4080 CTTTGAAGAACTTGT 4139  
 QY 1379 rPheArgGIYLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleG 1399

Db 4140 CTTCGGTGGCTTGCACACAGCGCTTCACTTCACTTGGCCCTTGTGTGTGTGTGTGTGTGT 4199  
 QY 1399 uArgAspGIYLeuLeuLeuTYrAsnGIYArgPheAsnGIYHisAspPheValAlaIle 1419  
 Db 4200 GCGTATACGGCTACTGT 4259  
 QY 1419 uGluValIleGlnGlnValGlnLeuThrPheSerAlaGIYGIYLeuSerThrThrVal 1439  
 Db 4260 CAGGTGTATCAGAAACAGGT 4319  
 QY 1439 lSerProPheValProGIYGIYValSerAspGIYGIYThrPheThrValGlnLeuLeuTY 1459  
 Db 4320 GTTCCGCTTGT 4379  
 QY 1459 rTYrAsnArgProLeuLeuGIYGIYThrGIYLeuProGlnIleProSerGIYGIYVal 1479  
 Db 4380 CTATATACAGCCCACTGT 4439  
 QY 1479 lAlaValAlaThrValAspGIYCysAspThrGIYValAlaLeuArgPheGlySerValIle 1499  
 Db 4440 AGCGGT 4499  
 QY 1499 uGIYAsnTYrSerCysAlaIleGlnGIYThrGlnGIYGIYSerIleYSerSerLeuAspLe 1519  
 Db 4500 GGGCACTACTCTGT 4559  
 QY 1519 uThrGIYProLeuLeuLeuGIYGIYValProAspLeuProGluSerPheProValArgMe 1539  
 Db 4560 GACAGGGCGCTGT 4619  
 QY 1539 rArgGlnPheValGIYCysMetArgAsnLeuGlnValAspSerArgHisIleLeuPheMetAl 1559  
 Db 4620 GCGGCACTTGT 4679  
 QY 1559 aAspPheIleAlaAsnAsnGIYThrValProGIYCysProAlaValHisAsnValCysAs 1579  
 Db 4680 CGACTTACTTCCAAACATGT 4739  
 QY 1579 pSerAsnThrCysHisAsnGIYGIYThrCysValAlaAsnGIYThrAspAlaPheSerCysGI 1599  
 Db 4740 CAGAGCATTTGT 4799  
 QY 1599 uCysProLeuGIYPheGIYGIYLeuSerCysAlaGlnGlnMetAlaAsnProGlnHisPhe 1619  
 Db 4800 GT 4859  
 QY 1619 eLeuGIYSerSerLeuValAlaThrPheGIYLeuSerLeuProIleSerGlnProTYr 1639  
 Db 4860 TCTGGGAGAGCGCTGT 4919  
 QY 1639 rLeuSerLeuMetPheArgThrArgGlnAlaAspGIYValLeuLeuGlnAlaIleThrAr 1659  
 Db 4920 CCTCACTCATGT 4979  
 QY 1659 gGIYArgSerThrIleThrLeuGlnLeuArgGlnGIYHisValMetLeuSerValGlnGI 1679  
 Db 4980 GGGGCGAGCAACATACCTGT 5039  
 QY 1679 yThrGIYLeuGlnAlaSerSerLeuArgLeuGlnProGIYArgAlaAsnAspGIYAspTY 1699  
 Db 5040 CACAGGGCTTCAAGCGCTGT 5099  
 QY 1699 PHisAlaGlnLeuAlaLeuGIYAlaSerGIYGIYProGIYHisAlaIleLeuSerPh 1719  
 Db 5100 GCATATATCAAGCTGT 5159  
 QY 1719 eAspTYrGIYGIYGlnArgAlaGlnGIYAsnLeuGIYProArgLeuHisGIYLeuHisIle 1739  
 Db 5160 CAACTATGGCAACAGACAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5219  
 QY 1739 uSerAsnIleThrValGIYIleProGIYProAlaGIYGIYValAlaArgGIYPheAr 1759

Dh 5220 GAGCAATATTACAGTCGGGGAGTTCTCGGGCCAGCCGAGAGTGGCCCTTCG 5279  
Qy 1759 GGIYCySeIeuGInGlyValArgValSerAspThrProGluGlyValAaSerLeuAspPr 1779  
Db 5280 GGGCGTTCGAGGGGTGTCAAGGTAAAGCAGACCTTGAAGGGGTTCACAGTCTAGATCC 5339  
Qy 1779 oSerHisGlyGluSerIleAaenValGluGlnGlyCySeSerLeuProAspProCyAspSe 1799  
Db 5340 CAGCCGCGGGAGAGCATCAATGTGGAGCCAGGCTGTAGCCCTGCAGACCCCTGTGACTC 5399  
Qy 1799 rAsnProCyProAlaAaenSerIYrCySeSerAsnAspTrpAspSerIYrSerCySeSerCy 1819  
Db 5400 GAATTCAGCCCTTACCAACAGCTACTACGCAATACCTGAAACAGCTATCTGTAGCTG 5459  
Qy 1819 sAspProGlyIYrIYrGlyYAspAsnCySeThrAsnValCyAspLeuAaenProCySeGluH1 1839  
Db 5460 TGTTCAGTACTAGTGTGACCACTGTACAAATGTGTGACCTGAAACCATGGAGACA 5519  
Qy 1839 gGlnSerValCySeThrArgIYrProSerAlaProHisGlyIYrThrCySeGluCySeProPr 1859  
Db 5520 CCAGTCCTGTGTACCCGAAACCCCAATACACCCACGGCTTACATCTGCGAGTGTTCAC 5579  
Qy 1859 oAsnIYrLeuGlyProIYrCySeGluThrArgIleAspGlnProCySeProArgIYrTrpTr 1879  
Db 5580 AAATTACTTGGGCCATATTGTGAGACCAAGATTGACCAACTTGGCCCGCTGGCTGTG 5639  
Qy 1879 pGlyHisProThrCySeGlyProCyAsnCyAspValSerIYrGlyPheAspProAspCy 1899  
Db 5640 GGGACACCCCAATGTGTGCATGCACTGTGATGTGACAAAGGCTTTCAGCCAGATGTG 5699  
Qy 1899 sAsnIYrHisSerIYrGluCySeHisCySeGlyGluAaenHisGlyIYrThrProProGlySeSerPr 1919  
Db 5700 CAACAGACCAAGTGGAGGTCCACCTGCAAGAGAAAGCATACCGGCCCGGCGAGCC 5759  
Qy 1919 oThrCySeLeuLeuCyAspCySeIYrProThrGlySeSerLeuSerArgValCyAspAspProG1 1939  
Db 5760 CACTGTCTCTTGTGTGACTGTACCCCACTGGTCTTGTGTCGAGCTGTGACCCAGCA 5819  
Qy 1939 uAspGlyGlnCySeProCySeIYrProGlyValIleGlyIYrArgIYrGlnCyAspArgCyAspAs 1959  
Db 5820 GGAAGGCGCAGTGTCCGTGCAAGCCGTGAGTGTGGGGCTGAGTGTGACCGCTGTGACAA 5879  
Qy 1959 nProPheAlaGluValThrThrAsnGlyCySeGluValAsnIYrThrAspSerCySeProArgAl 1979  
Db 5880 CCGTTTGTCTAGGTGACACCAACAAATGTGTGAAGTGAATTACCAACAGCTGCCACAGGCG 5939  
Qy 1979 aIleGluAlaGlyIleTrpIYrProArgThrArgPheGlyLeuProAlaAlaIaIaProCy 1999  
Db 5940 CATAGAGGCTGGGATCTGTGGCCCGGACCGGCTTCGGGCTACTGCTGCTGCCCTG 5999  
Qy 1999 sProIYrGlySerPheGlyIYrAlaValArgHisCyAspGlyH1sArgGlyIYrTrpLeuPr 2019  
Db 6000 CCCCAGAGGCTCTTGTGGACCTGTGTGGCCCACTGTGATGAGACAGGGGTGGCTGCC 6059  
Qy 2019 oProAsnLeuPheAaenCySeThrSerIleThrPheSerGluLeuIYrGlyPheAlaGluArg 2039  
Db 6060 CCAAAACCTTTCACATGCAAGTCACTCACTTCAGAACTAAAGGGCTTCGTGAGCG 6119  
Qy 2039 gLeuGlnArgAsnGluSerGlyLeuAspSerGlyIYrSerGlnGlnIleuAlaLeuLeu 2059  
Db 6120 GCTGCAAGAGAACAAATCGGCGCTGAGACTCAGAGCGCTCCAGAGGCTTACGCTGCTCT 6179  
Qy 2059 uArgAsnAlaThrGlnHisThrAlaGlyIYrPheGlySerAspValIYrValAlaIaIaIYrG1 2079  
Db 6180 GCGCAATGCCCAAGACACCTCTGGCTACTTCGGCAGTGTCAAGGTGGGCGCTTACCA 6239  
Qy 2079 nLeuAlaThrArgLeuLeuAlaHisGlyIYrThrAlaArgIYrPheGlyIYrLeuSerAlaIaTh 2099  
Db 6240 GCTGGCCACAGGCGCTGTGCTCAAGAGTCCCAAGCGGGGCTTGTGGGCTATCCGCGCAC 6299  
Qy 2099 rGlnAspValHisPheThrGluAaenLeuLeuArgValGlySerAlaLeuLeuAspThrAl 2119  
Db 6300 ACAAGAGTACACTTACCGAGGAATCTGCTGAGGGTGGGAGCGCCCTCTCGATGTGAGC 6359

Qy 2119 aAsnIYrArgHisPTrpGluLeuIleGlnGlnThrGluGlyGlyIYrThrAlaTrpLeuLeuG1 2139  
Db 6360 CAACAGAGGCACTGGGAACTGAACTCCAGAGACAGAGAGGTGGCACCGGCTGGCTGCCA 6419  
Qy 2139 nHisIYrGluAlaIYrAlaSerAlaLeuAlaGlnAsnMetArgHisPThrIYrLeuSerPr 2159  
Db 6420 GCATTAAGAGGCTTACCCAGCGCCCTTGGCCAGAAATATGGGCAACCTTACTTAAGGCC 6479  
Qy 2159 oPheThrIleValThrProAsnIleValIleSerValIYrArgLeuAspIYrGlyAspAsp 2179  
Db 6480 CTTACCAATGTGTACACCCCAATTTCAATCTCTGTAGTGGGAGCTGGAATTAAGGGAACTT 6539  
Qy 2179 eAlaGlyAlaIYrLeuProArgIYrGlyAlaIaLeuArgGlyGluGlnIleProProAspLeuG1 2199  
Db 6540 TGTGGAGACCAAGCTGCCCGTTATGAAAGCATTAAGTGGGAGAGGCCCGCCAGATGTGTA 6599  
Qy 2199 uThrThrValIleLeuProGluSerValPheArgGluThrProProValIaIaTrpAla 2219  
Db 6600 GACCAAGTCAATTTGCGAATACTGTGTCAAGAGATGCCCCCTCCATGTGTAGTGTGCG 6659  
Qy 2219 aGlyProGlyGlyAlaGlnGluProGluGluIleuAlaArgArgGluAlaArgHisPProG1 2239  
Db 6660 AGAAGCTGTGAAAGCCAGAGAGAGAGAGAGCTGGCCCGGCGAGCGAGACCCAGCA 6719  
Qy 2239 uLeuSerGlnGlyGluAlaValAlaIaSerValIleIleIYrArgThrIleuAlaGlyLeuLe 2259  
Db 6720 ACTGAGTCAGAGAGGCACTGGCGCAGTGTCAATCATCTACCAACCTTGGCTGAGTGTCT 6779  
Qy 2259 uProHisAsnIYrAspProAspIYrAspSerLeuAlaGlyValProIYrAspProIleIleAs 2279  
Db 6780 GCCCCAACAATTAAGACCCCATTAAGCGTACCTTGAAGTGTCCCAAGGCCCGCCGATCA 6839  
Qy 2279 nThrProValIaIaSerIleSerValHisAspAspGluGluIleuLeuProArgAlaIaLeuAs 2299  
Db 6840 CACCCCTGGGTAGATGACATCGCTTCAACAGACAGAGAACTCTGCGCGGGCTGTGGA 6899  
Qy 2299 pIYrProValIYrValGlnPheArgLeuLeuGluIYrGluGluIYrArgThrIYrAspProIleCy 2319  
Db 6900 CAAGCCAGTACCGTGTGAGTTCGGGCTGTGAGAGCCAGAGAGGAGCAAGCCATCTG 6959  
Qy 2319 sValPheTrpAsnHisSerIleLeuValSerGlyIYrThrGlyGlyIYrTrpSerAlaArgIYrCy 2339  
Db 6960 TGTCTTCTGGAACCATTAATCTGTGTGAGTGCACGTGGGCTGTGCTGCCCGAGGCTG 7019  
Qy 2339 sGluValIaIaPheArgAsnGluSerHisValSerCySeGlnCyAsnHisMetThrSerPh 2359  
Db 7020 CGAAGTGTCTTCCGTAAACAGAGCCACGTCACTGCGCACTGCAACCAATGTACAGCTT 7079  
Qy 2359 eAlaValIleuMetAspValSerArgArgGluAaenGlyGluIleuProLeuIYrThrIle 2379  
Db 7080 TGCATGTCTTAATGATATGTCAAGAGAGAGAGAGGAGATCTTGTGCCATGAAGACCT 7139  
Qy 2379 uThrIYrValAlaIaLeuGlyValIYrThrIleuAlaIaLeuLeuLeuThrPhePheHeLeuTh 2399  
Db 7140 GACATAGTGTGGCCCTTGTGAATCACTTGTGTGCTGCTGAGTCACTTCTCTTCTCCAC 7199  
Qy 2399 rLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIaIaLe 2419  
Db 7200 CTTCTCTCGGCGCTCCGCTCCCAACAGACAGGCACTCCAGCACTTCAAGCTGCTCT 7259  
Qy 2419 uGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAaenGlnAlaAspLeuProPheAlaCy 2439  
Db 7260 GGGGCTGGCCCAAGCTGTCTTCTCTGGGCAATCAACAGAGCTGAGACTCTCTTGTGCTG 7219  
Qy 2439 sThrValIleIleAlaIleLeuLeuHisPheLeuIYrLeuCySeThrPheSerTrpAlaIaLeu 2459  
Db 7320 TACAGTATGTGCTATCTGTGCAATTCCTGTACCTTGTCACTTCTCTGGGCTGCTGCT 7379  
Qy 2459 uGluAlaLeuHisPLeuIYrArgAlaLeuThrGluValArgAspValAsnThrGlyProMe 2479  
Db 7380 GGAAGGCTTACACTTATCCGGCGGCTCAAGAGGTGGGAGAGTCAATGCCAGTCCCAT 7439

2479 CARGHERYTYMCTLEUGLYTPGLYVALPROLAPHEIETHRGILYLAVALAIGI 2499  
 DB GCGTTCCTACTACAGCTGCGGCGGCGTCCCTCTTCATCAAGGCTCGCTGCGG 7499  
 2499 YLEUAPPROGUGLYTYRGLYANPROASPHECYETRPLEUSERTILEYTRASPETHLE 2519  
 DB CTGGATCCGAGAGGCTGAGGAGCCCTGACCTTCGCTGCTCTCTGTATGATACCT 7500  
 2519 UILETRPSEPHLEAGLYPROVALAPHEALAVASERMESERVALPHELEUTYRIL 2539  
 DB CATCGAGGCTTGTGCGGACGAGTGGCTTCTCTTCGAGGAGGCTTCTCTATAC 7560  
 2539 ELEUHALAALATGALASERCYVALAALAGLNGLYPHEGLYLYGLYGLYPROVA 2555  
 DB CTGTCGCGCCGAGGCTCTGTGCTGCGCCACGCGAGGCTTCGAGGAAGAGGCTGCT 7620  
 2559 ISEGLYLEUGLNGLYPROSEPHLEALVALLEUUEUUEUSERVALATHTRPLEUENAL 2579  
 DB CTGCGGCTGCGCTCTCTCTCTCAAGGCTCTCTGCTGCTGAGTGCACAGTGCCTGCTGCGC 7680  
 2579 ALEUUESEVALAANSEASPTHTLEUUEPHENISTYLEUENALATHRCYASNCY 2599  
 DB ACTGCTCTCTGTCACAGTGAACACTGCTCTTCCACTACCTCTGCTGCTGCAATTG 7740  
 2599 ATLEGLNGLYPROPHLEIPEHELEUSERTYVALYALLEUSERTYGLYVALAAGLYAAL 2619  
 DB TGTCCAGGCGCCCTTCATCTCTCTCAACGCTGTGCTGCTGAGCAAGAGTCCGGAAGC 7800  
 2619 ALEUYSLEUALACYSEASRGLYSPROSEPROASPPROALALAUTHTHTLYSESTH 2639  
 DB ACTCAAGTTCGCTGCGGAGGAGGAGCCGAGTCCGAGCTGCTGCTGCACTAATGATAC 7860  
 2639 ILEUHTSESERTYRASNCYSPROSEPROTYRILAPAGLYARGLEUTYGLNPROLY 2659  
 DB CCTGACCTGCTCTTAACCTGCGCCAGCCCTACGAGAGCGGAGGCTGATCAACCTTA 7920  
 2659 RGLYAPSERIALAGLYSERLEUENHISERTHTERARGSERGLYLYSERGLNPROSERTY 2679  
 DB CGGAAATTCGCTGCTGCTCTCAACAGTCCAGCCGCTCCGCAAGAGTCAACCCAGCTA 7980  
 2679 ILEPROPHLEUENALUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLY 2699  
 DB CATCCCTCTCTGCTGAGGAGGAGTCCAGCTGAGACCTGCGCCAGTTCCTCGGCT 8040  
 2699 UGLYAPPROGLYSERLEUENHISERTHTERARGSERGLYLYSERGLNPROSERTY 2719  
 DB AGGGAGCCCAAGCGGCTGCTCTGAGAGGCGCAAGCCCAACATGATCCGACACGGA 8100  
 2719 NSETPSERASPLEUENALUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLY 2739  
 DB CTCTACAGTGAACCTGCTCTGAGAGTGAACAGAGTGTCTCTATGCTCTCAACCTC 8160  
 2739 ISETPSERASPLEUENALUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLY 2759  
 DB ATCAGAGCAGC-----GAGGAGAGAGAGAGAGAGGCGCTTCCTCGGAGACA 8220  
 2759 NGLYTPAPSERLEUENGLYPROGLYVALAGLYARGLEUPROLEUENHISERTHTPROLY 2779  
 DB GGGCTGCGGACAGCTGCTGCTGCTGAGTGAAGTCAAGTCCCTTCACAGTACCCCA 8268  
 2779 ASAPGLYGLYPROGLYPROGLYLYVALAAPROTHPROGLYASPPHEGLYTHTHALY 2799  
 DB GATGAGGAGGCTGCGGAGTCTGCGGAGGCTCTGCGGAGGCTTGGAGCAACCAACAA 8328  
 2799 AGIUSERSERGLYAMGLYVALAAPROGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLY 2819  
 DB GAGAGATAGTGTAGTGGGCGCTTGAAGAGGCGCAAGGAGGAGTGAAGTGGCTTAC 8388  
 2819 TARGGLYGLYSERLEUENGLYPROLEUPROGLYSETERALAGLNPROHIALYGLYILE 2839  
 DB TCGGAGAGGCTCTGCGGAGCCCTTCGCGGCGCTTCAACCACTCAACAAAGGAGTCT 8448  
 2839 ULYSYLYSRYBLEUPROTHILSESGLYLYSSESESESESESESESESESESESESE 2859

DB 8508 CAAAGAAAGTGTCTGCCACCATCACAGGAAAGAGGCTCTTAAGGCTGCCCTGGA 8567  
 2859 UGLNCTHRCGLYSE 2879  
 DB GCAAGGACAGAGGCTCTCTGCGGCTCTCTCAACAGTGAAGGAGGCTGCGGACCTCC 8627  
 2879 OPROARGPROPROARGINSELEUGLUGLYUGLYUGLYUGLYUGLYUGLYUGLYUGLY 2899  
 DB TCCCGCCCACTCGCGCGCTCTGCAAGAAACACTGAATGGGCTCATGCTGCTGCGC 8628  
 2899 AMESERTILEYALAGLYTHRVALAAGGLUAPSERSESESESESESESESESESESESE 2919  
 DB CATAGAGATCAACCGAGCAGCGAGTGAAGTGAAGTCTCTGCTGCTCAATTCCTCTT 8688  
 2919 EASPHLEUENHIS 2923  
 DB 8748 TAACTCTGCTGAT 8760

## RESULT 7

D87469 8924 bp mRNA linear PRI 06-OCT-2001  
 LOCUS Human mRNA for KIAA0279 gene, partial cds.

DEFINITION D87469  
 ACCESSION D87469.1 GI:1665820  
 VERSION KIAA0279.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

## TITLE

Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayashi, Y.,  
 Ohara, O., Tanaka, A., Kotani, H., Miyajima, N., and Nomura, N.  
 Prediction of the coding sequences of unidentified human genes. VI.  
 The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 analysis of cDNA clones from cell line KG-1 and brain

JOURNAL DNA Res. 3 (5), 321-329 (1996)

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## 1. .8924

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="HA6133"

## /sex="male"

## /tissue\_type="brain"

## /clone\_lib="psport 1"

## 1. .8924

## /gene="KIAA0279"

## &lt;1. .7227

## /gene="KIAA0279"

## /note="Similar to D.melanogaster cadherin-related tumor

## suppressor"

## /codon\_start=1

## /protein\_id="BA13407.1"

## /db\_xref="GI:1665821"

## /translation="ESVPLGLVLAHQAIDADAGDNARLEYRLAGVGHDPPTINNGT

## GWSVLAELREERDPSFVKAARDHGTPTALTAASVTVLVLDVNDNPPFTQPEYTV

## RLNEDAVGTSVTVSAVDRAHSHVITYQTSGNTRRPFITSGGGLVSLAPLDY

## KLEGOYLAVTASGTRDPAOIYVNTDANTHRPFOSSHVTNVNEDBPAGTIVVL

## ISATDPTGNARITPTMEDSIPQFRIDADGATTPQARLDYEDVSVTLAATARDNG

## IPQSDPTTYELIYNDVNDNAPQFLRDSYGSYEDVPPPTSVTQISATRDSATNR

## VFTFGGDDGDFIVESTSGVTRILRLDRNVAQVLAIVADGMPAPKPMKV

## TVTVLDVNDPVPVEODEPVEENSPIGLAIVATATPDEBETNAQIOWIYEGNI

## PEVQOLDIFSGELTALVDLYBDRPVLVLAQTSAPLVSRATVHVALLRNDPPTV

## GNFELFNNTYTRSSSPGGAIGRVAPHPDISDSILYTFERGNELSLVILNASTGE

LKLSALDNNRPLEAIVMSVSDGVHVTQAQALRVIIIDEMLTHTSITLRLBDMSPER  
 RFLSEDLGLPIQAVATLTPDPDHYVNVQDRTDAPGHLINLSISGQPPGGGPP  
 PFLSEDLQERLYLNRSLTLTAISAQVLPFDNII CLRPENKNCVAVLEAFAP  
 IASSSVLPFRPHVGLRCRCPPGTGYCETBYDLCSYRGPGRGCRSREGYTCIL  
 CRDVTYGHCEVASASRGCTPGVCNKGCTCNLLVGRKCCPSDPEPKYQVTTLS  
 PPASPIITFRGLRORPHITLALSPATKRDGILLYNGRPNRKHDDVALEVIQEVOLT  
 PSAGSITTVSBPVGVSDQWMTVOLKYNKPLLGQTVLPQSGEBAKVAVTVYDGC  
 DTGVALARGLVANYSCAAGTQGGSKSLDTGLLGGVPLDPSPFPVAMRQVYGC  
 MRNLQVDSRHLDMADFIANNGTVPCCPAKKVCDNSNTCHNGGTVCNMDASCEPLG  
 FGKSCAQEMANPOHFLGSSLYVAMHGLPLISQPMYLIAMPTROADGVLLQATYRGR  
 STITOLAREGVMLSEVGTGLQASLRLEPRANDGMHHAOLGASGGGHILSP  
 DYGOORAGNGLPRHLGHLSNITGGYIIPRPGVYARPRGCLQCVTVSDTPEVNSL  
 DPSSGESINVRQCSLPDPCDSNCPNASCNDSDYS CSCDPGYEDNCTNCDNL  
 PCHEQSVCTKRKSAHGYTCRCPPNYLDPCEYTRIDQCPGMMGHPTGCGCNDVSK  
 GFDPDNCITSGECHKENHYRPPSGPTCLDCYPTGSLSHVCPEDQCCPKGVIG  
 ROCDCNDPFAVLTNGCEVNVYDCSPRALBGMWPRRPGIPAAAPPKASFGTAAR  
 HCDEHGMPLPMLFVCTSTITPSLTKGPRALORNSGQISDGRSQQALLANAOHTA  
 GYFSGDVYAVQOLATRLAHSTQKRFSLATODVHTFENLRVGSALLDTRAKHME  
 LIQOTBEGTAMLOHTEAVASALAQNMHRTYLSPTVTPNIVISVARDKGNFAGAK  
 LPRYALRGEQPPDLETVIILPESVRETPPVPRAGGEBOBEILARQRHPLS  
 OGEAVASVITRTLAGLPLPHYNDPKRLAPKRPINTPVVISVHDEBLPALD  
 KPVYQFPLETBERTEKPLCYEMNHSILVSGGWSARGCEVFNESHVSQCNHMT  
 SPAYMDVSRRENGILPLKTLTYVALGVTLLAALLPPELTILRIEENOHGRYRL  
 TLAAGLADVPLIGINQNDLPRACTVIAIILHPLYICTPSNALBALHTALTEVRD  
 VNTGPMRTYMLGMPVPAFITGLAVGLDPEGNDPFWLSIYDTLMSFAPVPAFV  
 SMSVPLXYILARASCAQORQEKXPGVGLQPSFVALLSDPMTALLSBNSTLL  
 FHYLPATCNICIQGPPITFLSYVYLSEKVRKALKACSRKPSDPAITLSTLSSNCP  
 SPYADGRLYOPDAGSLHSTSRGKQOPSTYPLILBBSALNGQOPGIDGPGSL  
 FLEQDQOHDHDDSDSLBNDQSGVASTHSDSREEREREREAAPRPGECQMS  
 ILRGRABERLPHASTPKDQGRPGKAPMFGDGTGTAKESGNGAPREERLRENGDLSRE  
 GSLGEPGSSAOPHKGILKCKLPTISKSSILRLPPLQCSGSSASBSGSGGPP  
 PRPPEPRLQLOQNGMPIMAMSIKAGTYEDBDSGSEPLFQFNFLH<sup>+</sup>

BASE COUNT 1727 a 2846 c 2504 g 1847 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 8924  
 Score: 12833.00 Matches: 2408  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 82.55% Indels: 0  
 DB: 9 Gaps: 0

US-09-916-849a-3 (1-2923) x D87469 (1-8924)

QY 516 GluservValProleuGlyTyrleuValleuHisValGlnAlaIleAspAlaAspAlaGly 535  
 DB 1 GAGAGCGTCCCTTAGGCTACCTGCTTCCATGTCGAGGCTATGACGCTGATCTGCT 60  
 QY 536 AspaenAlaArgLeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIle 555  
 DB 61 GACATATGCCCGCTGGAATACCGCCTTGCTGGGGTGGGACATGACTTCCCTTCAACATC 120  
 QY 556 AsnaenGlyThrGlyTyrPileserValAlaAlaGluLeuAspArgGluGluValAspPhe 575  
 DB 121 AACCATGACACAGGCTGATCTCTGTGGCTGTGAACCTGGACCGGAGAGAAAGTTATATTC 180  
 QY 576 TyrserPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSer 595  
 DB 181 TACACCTTTGGGGTGGAGCTCGAGACATGGCACTTCCAGCACTGACTGCTGGCCAGT 240  
 QY 596 ValserValThrValleuAspValAsnAspAsnaenProThrPheThrGlnProGluTyr 615  
 DB 241 GTCAAGCTGATCTGCTGATGTCAACGACATCCAACTTTAACCAACCAAGCAAGTAC 300  
 QY 616 ThrValArgLeuAsnGluAspAlaAlaValAlaGlyThrserValValThrValserAlaAla 635  
 DB 301 ACAAGTGGCTCAATGAGATGACGCTGTGGCACCAAGCGTGTACCGTGCACCTGTG 360  
 QY 636 AspaTgaAspAlaHiserValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArg 655  
 DB 361 GACCTGATGCTCATAGTCTCATCACTTACAGATCACCAAGTGGCAATCTCGAAGACCGC 420

QY 656 PheSerIleThrSerGlnserGlyGlyGlyLeuValSerleuAlaLeuProleuAspTyr 675  
 DB TTTCTTCATACACACCAAGGAGGTGGGTGGTATCTCTTGCCTGCGCATGAGTAC 480  
 QY 676 LysleuGluArgGlnTyrValleuAlaValThrAlaSerAspGlyThrArgGlnAspThr 695  
 DB 481 AAATCTTGAAGCGCGAGTATGTTGGTGGCTTAAACCCCTCCGATGACACTCGGACAGACG 540  
 QY 696 AlaGlnIleValAlaAsnValThrAspAlaAsnThrHisAspProValPheGlnserSer 715  
 DB 541 GCACAGATTGGGTGAATGTCACCGACCGCACCAACCCATGCTCTTCTTCAAGCTCC 600  
 QY 716 HisTyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValleuIleSer 735  
 DB 601 CACTATACAGTGAATGTTAATGAGACCGGCGGACAGCACACAGGTGTCTGATGACG 660  
 QY 736 AlaThrAspGluAspThrGlyValAsnAlaArgIleLeuThrTyrPheMetGluAspSerIle 755  
 DB 661 GCCACGATGAGACACACAGGTGAATGCCCGCATACCTTCAATGAGAGACAGATC 720  
 QY 756 ProGlnPheArgIleAspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyr 775  
 DB 721 CCCAGTTCGCGCATGATGACAGACACGGGGCTGTACACCCAGGCTGAGCTGACTAC 780  
 QY 776 GluAspGlnValSerTyrThrleuAlaIleThrAlaArgAspAsnGlyIleProGlnLys 795  
 DB 781 GAAGACCAAGGTCTTACACCTTGGCCATTAATCTCTCGGACCAATGACATTCGCCAAG 840  
 QY 796 SerAspThrThrTyrleuGluIleleuValAsnAspValAsnAspAlaAlaProGlnPhe 815  
 DB 841 TCCACACACCATACCTGAGATCTCTGTGAACAACATGATGACATGACATGCCCCACAGTTC 900  
 QY 816 LeuArgAspSerTyrGlnGlySerValTyrGlyLysAspValProProPheThrSerValleu 835  
 DB 901 CTGCGAATCTCTACAGGCGCATGCTATGAGATGTGCCACCTTCACTAGGTCTCTG 960  
 QY 836 GlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGln 855  
 DB 961 CAGATCTCAGCACTGATCTGATCTGGAATCTTAATGAGAGGGGTCTTCAACCTTCAA 1020  
 QY 856 GlyGlyLysAspAspGlyAspAspPheIleValGlnSerThrSerGlyIleValArgThr 875  
 DB 1021 GGAAGCCACATGAGAGCGGTGACTTATGTTAGTCAAGTCAAGGATGATGTCGAAACG 1080  
 QY 876 LeuArgArgLeuAspArgGluAsnValAlaGlnTyrValleuArgAlaTyrAlaValAsp 895  
 DB 1081 CTACGAGAGCTGATTCAGAGAGACGTGGCCACAGTATGCTTGGCGGACATAGCTGAGAC 1140  
 QY 896 LysGlyMetProProAlaArgThrPrometGluValThrValThrValleuAspValAsn 915  
 DB 1141 AAGGGGATGCCCCCAGACCCGACACCTATGGAAGTGAACGTCACTGTGTGGAATGTAAT 1200  
 QY 916 AspaenProProValPheGluGlnAspGluPheAspValPheValGluLysAspSerPro 935  
 DB 1201 GACATATCCCCCTGCTTGGAGCAGATGAGTTGATGTGTTGGAGAAAGAACAGCCCC 1260  
 QY 936 IleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGln 955  
 DB 1261 ATTGGGCTAGCGGTGGCCCGGGGTCAAGCCATGACCCCATGAAAGCAACAAAGCCAG 1320  
 QY 956 IleMetTyrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSer 975  
 DB 1321 ATTATGATCCAGATTGGAGGGGCAATCATCTGAGGTCTTCCAGCTGAGACATCTTCTCC 1380  
 QY 976 GlyGluLeuThrAlaLeuValAspLeuAspThrGlyLysAspArgProGluTyrValleuVal 995  
 DB 1381 GGGAGGCTGACAGCCCTGAGTACCTTAACATACAGAGACGGGCTGAGATGATGCTCTGCTC 1440  
 QY 996 IleGlnAlaThrSerAlaProleuValSerArgAlaThrValHisValArgLeuLeuAsp 1015  
 DB 1441 ATCCAGGCGACATGATGCTCTCTGTGAGCGGGCTTCAAGTCCAGTCCGCTCTTGAC 1500  
 QY 1016 ArgAsnAspAsnProProValleuGlyAsnPheGluIleLeuPheAsnAsnTyrValThr 1035

Db 1501 CGCAAGACACACCACAGTGGGCACTTGAAGTCTTTCAACAACTATGACCC 1560  
 Qy 1036 AenAArgSerSerPheProGlyValAlaIleGlyArgValProIleHisAapProAap 1055  
 Db 1561 AATCGCTCAAGACGCTTCCCTGGGGGTGCATTGGCCGAGTACTCTGCCATGACCTTGAT 1620  
 Qy 1056 IleSerAapSerLeuThrTySerPheGluArgGlyAsnGluLeuSerLeuValIleLeu 1075  
 Db 1621 ATCTCAGATAGTGTGACTTACAGCTTTGAGCGGGGAATATGACTCAGCCGTGGTCTGCTC 1680  
 Qy 1076 AsnAlaSerThrGlyGluLeuGlyLeuSerArgAlaLeuAapAanAanArgProLeuGlu 1095  
 Db 1681 AAGGCTCCACCGGTAGCTGAAGCTAAGCCGCGCACTGGACAACACCGGCTCTGGAG 1740  
 Qy 1096 AlaIleMetSerValIleuValSerAapGlyValHisSerValThrAlaGlnCysAlaLeu 1115  
 Db 1741 GCCATCATGAGCTGTGGTGTGATGAGACGGGTACACAGCGTGAACCGCCAGTCCGCGCTG 1800  
 Qy 1116 ArgValThrIleIleThrAapGluMetLeuThrHisSerIleThrLeuArgLeuGluAap 1135  
 Db 1801 CGGTGACCATCATCAACCATGATGATGCTCAACCAAGCATCAAGCTGGCGCTGGAGGAGC 1860  
 Qy 1136 MetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAla 1155  
 Db 1861 ATGTCAACCGAGGCGCTTCTGTCAACACTGTAGGCTCTTCATCAAGCGGTGGCGCGCC 1920  
 Qy 1156 ThrLeuAlaThrProAapPheHisValValIlePheAanValGlnArgAapThrAapAla 1175  
 Db 1921 AGCTGGCCACGCCACCGACACAGTGTGTGTCTTCAACCTTACAGGGAGACACCAAGCC 1980  
 Qy 1176 ProGlyGlyHisIleLeuAanValSerLeuSerValGlyLeuProProGlyProGlyGly 1195  
 Db 1981 CCGGGGGGCGCACTCTCAACGTGAGCGTGTGGTGGGCGACGCGCGCGCGGGGGG 2040  
 Qy 1196 GlyProProPheLeuProSerGluAapLeuGlnGluArgLeuTyLeuAanArgSerLeu 1215  
 Db 2041 GGGCGCCCTCTCTGCTCTGAGGACCTGCAGAGACGCTTATACCTCAACCCAGACCTG 2100  
 Qy 1216 LeuThrAlaIleSerAlaGlnArgValIleuProPheAapAapAanIleCysLeuArgGlu 1235  
 Db 2101 CTGACGGCCATCTCGGCAACGCCGCTGTGCTCTTGAACACAACTCTGCTCGGGAG 2160  
 Qy 1236 ProCysGluAanTyMetArgCysValSerValIleuArgPheAapSerSerAlaProPhe 1255  
 Db 2161 CCTTGGAGAACTACATGCGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220  
 Qy 1256 IleAlaSerSerSerValIleuPheArgProIleHisProValGlyGlyLeuArgCysArg 1275  
 Db 2221 ATGCGCTCTCTCTCGGTGCTCTTCCGACCATCACCCGTCGAGGGCTGCGCTGCC 2280  
 Qy 1276 CysProProGlyPheThrGlyAapTyArgValIleuThrGlyValAapLeuCysTySerArg 1295  
 Db 2281 TGGCCGCCCGGCTTCAAGGGTACTACTGCGAAGACCGAGGTGACCTCTGCTACTGCGG 2340  
 Qy 1296 ProCysGlyProHisGlyArgCysArgSerArgGluGlyGlyTyThrCysLeuCysArg 1315  
 Db 2341 CCTGTGGGCGCCACCGGCGCTGCCGACGCGGAGGGCGGCTACACTGCTGCTGCTGCT 2400  
 Qy 1316 AapGlyTyThrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGly 1335  
 Db 2401 GATGGCTACACGGGTAGACACTGTAGGTGAGTGTGCTGCTCAAGCGGTGACACCCGGGT 2460  
 Qy 1336 ValCysLeuAanGlyGlyTyArgCysValAanLeuLeuValGlyGlyPheLeuCysAapCys 1355  
 Db 2461 GTCTGAGAGATGGGGGCACTGTGTCAACTGCTGTGTGGCGGTTCAAGTGCATTGC 2520  
 Qy 1356 ProSerGlyAapPheGluArgProTyArgGlnValThrThrArgSerPheProAlaHis 1375  
 Db 2521 CCATCTGGAGACTTTCAGAGCCCTTACTGCAAGTATCAACACGCGAGCTTCCCGGCCAC 2580  
 Qy 1376 SerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrIleuAlaLeuSerPhe 1395

Db 2581 TCCCTCATCACTTTTGGCGGCTGGCGCAAGCGTTTCACTTCAACCCGCGCTCTGCTT 2640  
 Qy 1396 AlaThrTyGluArgAapGlyLeuLeuLeuTyArgGlnArgPheAanGluTyHisAap 1415  
 Db 2641 GCCACAAAGAGACGGGACGGGTGCTGTGTTCATATGGCCGTTTCAATGAGAAAGATGAC 2700  
 Qy 1416 PheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyLeuSer 1435  
 Db 2701 TTTGTGGCCCTCAAGGTATCAAGAGACAGGTCAGCTCACTTCTCTGCAAGGGAGTCA 2760  
 Qy 1436 ThrThrThrValSerProPheValProGlyGlyValIleSerAapGlyGlnThrPheThrVal 1455  
 Db 2761 ACCACCAAGGTGTCCCATTCGTGCGCGAGAGATCAATGTATGTGCCAAGTGCATCGGTG 2820  
 Qy 1456 GlnLeuTyTyTyTyAanTyPheLeuLeuLeuGlyGlnThrGlyLeuProGlnTyProSer 1475  
 Db 2821 CAGCTGAATACTTACATTAAGCCACTTGTGGTCAAGACGGGCTCCACAGGGGCCATCA 2880  
 Qy 1476 GlnGlnTyValAlaValValThrValAapGlyCysAapThrGlyValAlaLeuArgPhe 1495  
 Db 2881 GAGCAGAAAGTGGCTGTGTGTGACGTGTGATGGCTGTGACACAGAGTGGCTTGGCTTC 2940  
 Qy 1496 GlySerValLeuGlyAanTySerCysAlaAlaGlnGlyTyThrGlnGlyGlySerTyHis 1515  
 Db 2941 GGATCTGTCTGGGCACTACTCTGTGTCTGCCAGGGCACCCAGAGTGGCAGCAAGAG 3000  
 Qy 1516 SerLeuAapLeuThrGlyProLeuLeuGlyGlyValProAapLeuProGlyLeuSerPhe 1535  
 Db 3001 TCTGTGATCTGACGGGGCCCTGTCTTATGGCGGGGTGCTGACTGCTGCCAGAGCTTC 3060  
 Qy 1536 ProValArgMetArgGlnPheValGlyCysMetArgAanLeuGlnValAapSerArgHis 1555  
 Db 3061 CCAGTCCCAATGGCGCACTTGTGGGTGACATGCGAATCTGCAAGTGAACAGCGCGCAC 3120  
 Qy 1556 IleAapMetAlaAapPheIleAlaAanAanGlyThrValProGlyCysProAlaTyHis 1575  
 Db 3121 ATAGACATGGCTGACTTATGTCACAAATGACACCGTGGCTGCTGCTGCTGCTGCTGCTG 3180  
 Qy 1576 AenValCysAapSerAanTyCysHisAanGlyGlyTyThrCysValAanGlnThrAapAla 1595  
 Db 3181 AACGTGTGACAGCAACTTGTCCACATGGGGGCACTTGTGTGAACAAGTGGGACCG 3240  
 Qy 1596 PheSerCysGlyCysProLeuGlyPheGlyGlyTySerCysAlaGlnGluMetAlaAan 1615  
 Db 3241 TTCAGCTCGAGTGGCCCCCTGGCTTGGGGCAAGACTGCGCCACAGAAATGCGCAAT 3300  
 Qy 1616 ProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSer 1635  
 Db 3301 CCACAGCACTTCTGGGACAGACTGTGTGTGTGGCAATGGCTTCTGCTGCCATCTTC 3360  
 Qy 1636 GlnProTyTyTyLeuSerLeuMetPheArgThrArgGlnAapGlyValIleuGln 1655  
 Db 3361 CAACCTGTGATCTCAAGCTTATGTTCCGACAGCGCCAGGCGAGGTGCTGCTGTCGAG 3420  
 Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeu 1675  
 Db 3421 GCCATCAACAGGGGGCGGACGACCATCACTTACAGCTTACAGAGGGCGACGTGATGTG 3480  
 Qy 1676 SerValGluGlyTyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAan 1695  
 Db 3481 AGCGTGAAGGACAGGGCTTCAAGGCTCTCTCTCTGCTGAGACCGAGCGGGCCAAAT 3540  
 Qy 1696 AapGlyAapThrHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAla 1715  
 Db 3541 GACGGTGACTGGCACTAGCAAGCTGCACTGGGAGCGAGGGGGGCGCTGGCATGCC 3600  
 Qy 1716 IleLeuSerPheAapTyGlyGlnGlnArgAlaGluGlyAanLeuGlyProArgLeuHis 1735  
 Db 3601 ATTCTGTCTTCTGATTATGGGACAGAGAGAGAGGAGCAACTGAGGCCCCCGGCTGAT 3660  
 Qy 1736 GlyLeuHisLeuSerAanIleThrValGlyGlyTyIleProGlyProAlaGlyGlyValAla 1755  
 Db 3661 GGTCTGACACTGAGCAATTAACGTGGCGGGAATCTTGGCGCACCGGGCGGTGTGACC 3720



QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGluGlyValAlaAn 1775  
 Db 3721 CCGGGCTTCGGGGGCTGTTTGCAAGGGTGTGGGGGTGAGGGAATACCGCAAGGGGGGTAAAC 3780  
 QY 1776 SerLeuAspProSerHisGlyGlySerHisLeuValGluGlnGlyCysSerLeuProAsp 1795  
 Db 3781 AGCTGGAAATCCCAAGCATGGAGGAGACATCAAGTGAGCAAGAGCTGTAGCCCTGGTAC 3840  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyr 1815  
 Db 3841 CCTTGGATCTCAACCCGCTGTCTGCTTACAGCTATTGGAGCAACATGGAGCAACTTAC 3900  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 Db 3901 TCGTGACGCTGATCATGAGTTACTATGGTGACACTGTACTATATGTGTGTGACCTGAC 3960  
 QY 1836 ProCysGluHisGlnSerValCysThrArgGlyAspSerLeuProHisGlyTyrThrCys 1855  
 Db 3961 CCGTGAGACACAGTGTGTGTACCCGCAAGCCAGTGCCCCCATGGCTATATCTTGC 4020  
 QY 1856 GluCysProProAsnTyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysPro 1875  
 Db 4021 GAGTGCCCCCAATTAATCTTGGCCATCTGTGAGACCAAGATTACAGCTTGTCCC 4080  
 QY 1876 ArgGlyTyrTrpGlyHisProThrCysGlyProCysAsnCysAspValSerIysGlyPhe 1895  
 Db 4081 CCGTGCTGGGGGACATGCCACATGTGGCCCATGCACTGTGATGTGACAAAGGCTTT 4140  
 QY 1896 AspProAspCysAsnIleThrSerGlyGluCysHisCysValGluAsnHisGlyThrPro 1915  
 Db 4141 GACCCAGACTGCAACAGCAAGCGCGAGTGCCATGCAAGAGAACCACTACCGGCC 4200  
 QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 Db 4201 CCAAGGAGCCCACTCGCTCTTGTGTACTGTCTACAGGCTCTTGTGCCAGATC 4260  
 QY 1936 CysAspProGluAspGlyGlnCysProCysIysProGlyValIleGlyValArgIleCysAsp 1955  
 Db 4261 TGTGACCCCGAGAGATGGCAGATGTCATGCAAGCCAGTGTCTATGGGGGTCAAGTGTAC 4320  
 QY 1956 ArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGluValAlaAsnTyrAspSer 1975  
 Db 4321 CCGTGAGAACCCCTTTGTCTAGAGTCAACCAAGGCTGTGAAGTGAATTAATACAGCC 4380  
 QY 1976 CysProArgAlaIleGlyAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAla 1995  
 Db 4381 TCCCAACAGAGCATTTAGGCTGGATCTGTGGCCCCGATCCGCTTCGGGGTGGCTGCT 4440  
 QY 1996 AlaAlaProCysProIysGlySerPheGlyThrAlaValArgHisCysAspGlyHisArg 2015  
 Db 4441 GCTGCTCCCTGTCCCAAGAGCTCTTTGGAGCTGCTGCGCCACTGTGATAGACACAGG 4500  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGluLeuIysGly 2035  
 Db 4501 GGGTGGCTCCCCCAACCTTTCACATGCAAGTCACTCATCCTTCAGAACTGAAGGAC 4560  
 QY 2036 PheAlaGluArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnIleu 2055  
 Db 4561 TTCCGTGAGCGGCTACAGCGGATAGTCAAGGCTTAGACTCAAGGCGCTCCCAAGAGCTA 4620  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisGlnAlaGlyTyrPheGlySerAspValIys 2075  
 Db 4621 GCCCTGCTCTCCGCAACCGCAAGCAGCACAGCTGGCTACTTCCGCGACCACTCAAG 4680  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 Db 4681 GTGGGCTTACAGCTGGCCAGCGGCTGTGTGCCACAGAGGACCCAGCGGGGCTTTGGG 4740  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeu 2115  
 Db 4741 CTGTGTGACCAACAGGAGCTGCACTTCACTGAGAAATCTGTGCGGGTGGGACGCCCTTC 4800

QY 2116 LeuAspThrAlaAsnIysArgHisGlyTrpGluLeuIleGlnGlnThrGluGlyValThrAla 2135  
 Db 4801 CTGAGCACAGGCCCAACAGCGGACTGGAGGCTGATCCAGAGACAGAGGGTGGACCCGC 4860  
 QY 2136 TrpLeuLeuGlnHisTyrGluAlaTyrAlaSerAlaLeuAlaGlnAspMetArgHisGln 2155  
 Db 4861 TGGCTGCTCCAGCACTATGAGGCGCTTACCGCAGTGGCCCTGGGCCCAAGATGGGCAACC 4920  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValIleArgLeuAsp 2175  
 Db 4921 TACCTAAAGCCCTTTCACATGCTCAAGCCCAACATTTGATCTCCGTAGTGGCTTGGAC 4980  
 QY 2176 LysGlyAspPheAlaGlyValAlaIysLeuProArgTyrGluAlaLeuArgGlyGluGlnPro 2195  
 Db 4981 AAAAGGAATTTGCTGGGGCCAAAGCTGGCCCGCTTACAGAGGCCCTGGCGGGAGAGCC 5040  
 QY 2196 ProAspLeuGlnThrThrValIleLeuProGlySerValPheArgGluThrProProVal 2215  
 Db 5041 CCGAGCTTGAGAACAAAGTCAATCTGTGAGTCTGTCTTCAAGAGACGCCCCCGTGTG 5100  
 QY 2216 ValArgProAlaGlyProGlyGlyValAlaGlnGluProGluGluLeuAlaArgArgGlnArg 2235  
 Db 5101 GTCAAGCCCCGAGGCCCGGAGAGGCCCAAGAGCCAGAGAGCTGGCACGGGCAAGGCA 5160  
 QY 2236 ArgHisProGluLeuSerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeu 2255  
 Db 5161 CGGCACCGGAGCTGAGGCAAGGTGAGGCTGTGGCCAGCGTCACTTACCGGACCCCTG 5220  
 QY 2256 AlaGlyLeuLeuProHisAsnTyrAspProAspIysArgSerLeuAspValProIysArg 2275  
 Db 5221 GCGGGCTTACGTGCTCATTAATGATGACCTTGACAAAGGCAAGCTTGGAGTCCCAAGAGC 5280  
 QY 2276 ProIleIleAsnThrProValIleSerIleSerValHisAspAspGluGluLeuLeuPro 2295  
 Db 5281 CCGATCATCAACACACCCGTGTGATGATCAAGCTTCATATGATGAGAGCTTCTGCC 5340  
 QY 2296 ArgAlaLeuAspIysProValThrValGlnPheArgLeuLeuGlnThrGluGluArgThr 2315  
 Db 5341 CGGGCCCTGGAACAAACCGGTCAAGGTGCAATTCGCGCTGGAGAGACAGAGAGGGAGC 5400  
 QY 2316 LysProIleCysValPheThrAsnHisSerIleLeuValSerGlyTyrThrGlyTyrSer 2335  
 Db 5401 AAGCCCATCTGTGCTCTTGAGAACCATTAATCTGTGTCAATGTGCACAGGTGGTGTG 5460  
 QY 2336 AlaArgGlyCysGluValIlePheArgAsnGlySerHisValSerCysGlnCysAsnHis 2355  
 Db 5461 GCCAGAGCTGTGAAGTGTCTTCCGCAATGAGAGCCACGCTCACTGCACTGCAACAC 5520  
 QY 2356 MetThrSerPheAlaValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuPro 2375  
 Db 5521 ATGACGAGCTTCGCTGTGCTCATGACGTTTCTGGCGGAGAAATGGGGAGATCTTGGCA 5580  
 QY 2376 LeuIysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhe 2395  
 Db 5581 CTGAAGACATGACATACGTGGCTCTAGGTGTCACTTGGGTGCTCTTGTGCTCACTTC 5640  
 QY 2396 PhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeu 2415  
 Db 5641 TTCTTCTCACTCTCTTGGGATCTGCGATCTGCGCTCAACCAACAGCGCACTCCAGCTTCC 5700  
 QY 2416 ThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeu 2435  
 Db 5701 ACAGCTGCGCGGGCCGTGGCTCAAGCTGATCTTCTCTCGGGAAATCAACAGGGCTGACCTC 5760  
 QY 2436 ProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSer 2455  
 Db 5761 CTTTGTGCTGCAAGTCAATGCTTCCATCTGTGCACTTCTGTACCTTGTCACTTTTCC 5820  
 QY 2456 TrpAlaLeuLeuGluAlaLeuHisIleLeuTyrArgAlaLeuThrGluValArgAspValAsn 2475  
 Db 5821 TGGGCTGCTGAGAGGCTTGGACCTGTATCCGGGCACTCACTGAGGTGGGAGTGTACAC 5880  
 QY 2476 ThrIysProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGly 2495



Db	5881	ACGGGCCCCAATGGGCTTCTACTACATGCTGGGCTGGGGCCGTGGCTGCTTCATACAGAGG	5940
Qy	2496	LeuAlaValGlyLeuAspProGluGlyTyrGlyValAspProAspPheCysTrpLeuSerIle	2515
Db	5941	CTAACCCGTGGAGCCCTGGAGCCCGAGAGGGCTCAGAGAACCTGACTTCTGCTGTGCTCTTCATC	6000
Qy	2516	TyrAspThrLeuIleTyrPserPheAlaGlyProValAlaPheAlaValSerMetSerVal	2535
Db	6001	TATACACAGCTCATCTGGAGATTTCGTGGCCGGGTGGCTTTCGCGCTCTGCATAGATGTC	6060
Qy	2536	PheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluIys	2555
Db	6061	TTCCCTGTACATCTCCGGCGGCCCGGGGCTCTGTGCTGCGCCAGCGGAGGGGCTTTGAGAG	6120
Qy	2556	LeuGlyProAlaSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThr	2575
Db	6121	AAAGGTCTCTGTGGGCTGGAGCCCTCTTCGCGGCTTCCTGCTGCTGAGGCGACG	6180
Qy	2576	TrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAla	2595
Db	6181	TGGCTGTGGGACTGCTCTCTGTCAACAGGAGAACCCCTTCCTTCACATACCTTTTGCT	6240
Qy	2596	ThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValIleLeuSerGlyIle	2615
Db	6241	ACCGCAATTGCATCCAGGGGCCCTTCATCTTCTCTCATATGGTGCTTAGACAGAG	6300
Qy	2616	ValArgIysAlaLeuIysLeuAlaCysSerArgIysProSerProAspProAlaLeuThr	2635
Db	6301	GTCGGAAAGCACTCAAGCTTGGCTCTGCAGCGGCAAGCCAGCCCTGACCTGCTCTGACC	6360
Qy	2636	ThrIysSerThrIleuThrSerSerTyrAsnCysAspSerProTyrAlaAspGlyArgLeu	2655
Db	6361	ACCAAGTCCACCTCGACCTCGCTCTACAACTCCCGGACCCCTCAGCAGAGAGGGCGGCTG	6420
Qy	2656	TyrGlnProTyrGlyAspSerAlaGlySerIleuHisSerThrIysArgSerGlyIysSer	2675
Db	6421	TACCAAGCCTTAGGAGACTCGGGCGGCTCTGTGCACAGCACAGTGGCTGGGCAAGGT	6480
Qy	2676	GlnAspSerTyrIleProPheLeuLeuArgGluIleSerAlaLeuAspProGlyGlnGly	2695
Db	6481	CAGCCCAAGCTCATCCCTCTTCTGTGAGGAGAGTCCCGCATGAAACCTGGGCAAGGG	6540
Qy	2696	ProProGlyLeuGlyAspProGlyIserIleuPheLeuGlyGlyIleAspGlnGlnHisAsp	2715
Db	6541	CCCCCTGGCCCTGGGAGATCCAGGAGCGCTGTCTTGGAAGGTCAAGACCAAGACATGAT	6600
Qy	2716	ProAspThrAspSerAspSerAspLeuSerLeuGluIleAspAspGlnIserGlySerTyrAla	2735
Db	6601	CTTACACAGCACTCGACAGTGACTGTGCTTTAAGAACACACAGAGTGGCTCTTAATCC	6660
Qy	2736	SerThrHisSerSerAspSerAspSerGluGluGluGluGluGluGluGluGluAlaPhe	2755
Db	6661	TCTACCCACTCATATCAGACATGAGAGAGGAAAGAGAGGAGAGAGAGAGGCGGCTTC	6720
Qy	2756	ProGlyGluGlnGlyTyrPAspSerLeuLeuGlyProGlyAlaGlyIleAspGlyProLeuHis	2775
Db	6721	CTTGAGAGAGCGGGCTGGATAGCTGTGCGGGGCTTGAGAGACAGAGACTCCCTTGAC	6780
Qy	2776	SerThrProIysAspGlyGlyProGlyProGlyIysAlaProTyrProGlyIysAspPheGly	2795
Db	6781	AGTACTCCCAAGATGGGGGCGCCAGAGGCGCTGGCAAGGCCCTCGGCGCAGAGACTTTGGG	6840
Qy	2796	ThrThrAlaIysGlyIserSerGlyAsnGlyAlaProGluGluIleAspGlyLeuArgGly	2815
Db	6841	ACCAACAGAAAGAGATGATGTCAGACGGGGCCCTTAGAGCGAGCTGGGAGAAATGGA	6900
Qy	2816	AspAlaLeuSerAspArgGluGlySerIleuGlyProLeuProGlyIserSerAlaGlnProHis	2835
Db	6901	GATGCCCTGTCTCGAGAGGGGATCCCTAAGGCCCTTCCAGGCTCTTTCGCCACAGCTCAC	6960
Qy	2836	LysGlyIleLeuIysIysIysCysLeuProThrIleSerGlyIysSerSerLeuLeuArg	2855

[illegible]

DNMRPLEAMSVLVDGHSVTAQCAKVTITITDEMLTHSITLRLIEDMSPERFESPLL  
 GLEFQAVATATPPDHVVENVQDRTBAAGHILNVSLVQAPRGPGGPEPLFSELD  
 LQERLYLNRSLLTALSAQVLPFDNDNICREPCENMYCVSLTARSDSAPFASSVL  
 FRPIHVGRLCRCPGEGFDYCEVETGLCYSRPCGPHRGCRSGRTGLCRDQYTG  
 EHCISVARSGRCTPGVCKNGGTCAVLVVGFKDCPSDPEKPYCOVATRSAGSST  
 TTRSPRORPHFTIALSPATKREDGLLVNGRFNEGDDVALAVETIOEQVLTSSGSEST  
 TTVSPFVRGVSQDGMHTVQLKTNKPLDGTGLFGQSPBOKVAVTVQCDTGVALLR  
 FESVLGANSYCAAGTQGSKSLDLTGLLLAGVPLDPSBPVRNRQVPGCKRNQVD  
 SRHIMADPFIANGTVPGCPACKVNCSTNCGGTCAVOMDAPCECPGLGSGSCA  
 QEMANPOHPLAGSLVAMHGLSLPISQPVLSLMPRTROADGVLLQALTRGSESTLLOL  
 REGHYMLSVETGLOASSLRLEPGRANDMHAQALGASGPGHAILSPYGGORA  
 EGNGLPRLGLHLSTVIGGIPGPAQVARGRGLQCTVRSUTPREGNLSIDPSHGES  
 INVEGCSLPDPCDSNPPCANYSCTNDMSISCSCEPVTGNCNTVCDLNPCEHQS  
 CTRKSPHGYTCECPVNLGPECETRIIDQPCRWMCMPHPTGCPNCNCSVSGKPDCCN  
 KTSBGCKKENHNRPPGSPYCLCDQCYPTGSLSRVCDDEDCQCPKCPVIGROCDRC  
 NLPALVTNGCEVNDSCPRALGAIWMPRTGFPALAPCPKSGFVAVHRCDEHSG  
 WLPFVNTGSIITFSELGPAERLQNRSGLDGSRSQLALLRNATORTGYSRSDV  
 KVAVOLARILAHSESTORGRLSATODPHENILNRNSALDTPKRWHELIOQTEG  
 GTANLQHTBEAVASLADQMRTTYSPTTITVNTVISVRLDKNRFAGALPRTYAL  
 RGEOPDLLETVVILPESVFRETPPVVRPAPGPAQSPBELARORRHPBELQGEAVAS  
 VLIYRTLAGLPHNDPDKRSILVRKPIINTPVVISVHDBELLPALDKPVVOC  
 RLLETERTKPICVFMNHSILVSGTGSMSAGCEVFPNESHVSQCNHMTSPAVLMD  
 VSRRENGELPLKLTLYVALGVTLAALLTPPLFLILIRSNHGIERBNJALGLA  
 OLVELLGINOADLPACTVIALILHPLVLCPTSMALLRALHLRYLTVRVYNTGPKR  
 PTVMLGKVPAPITGLANGDPBRYGANDFCWLSITDILMSFACPVAFAVMSVFLY  
 ILAARASCAADQOGFERKGPVSGLOPSFAVILLSATWLLLSVNSDTLLPHLIFAT  
 CNCIQGPPIFLSYVLSKEVRKALCLASRKRPDPALTTSTNCSPYADGR  
 LYOPYGDASGLSHSTRSKQSPYIPLLEBESALNPGQSLGSGSLFLRQDQ  
 OHDPPTDSDSLHSTEDQSGYASTYSIDSEBEEBEEBEPAPBEOQMDSLFLEGQAE  
 RLPHSTKDGSGPGPKAPMPGDPRTGTAKRESGNGCAPEPRARZENDALSRGSGIPLP  
 GSSAQPHKGIKKKCLKPTISKSSILRPLEDCTSSGSSASBSRSGRGPFRPPRPPRO  
 SLOBQNKVMPHMSIKAGTDESSSEPLFPNPLH"

BASE COUNT 4543 a 7241 c 7458 g 5128 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 24370  
 Score: 12339.00 Matches: 2907  
 Percent Similarity: 36.92% Conservative: 0  
 Best Local Similarity: 36.92% Mismatches: 8  
 Query Match: 79.38% Indels: 4967  
 DB: 6 Gaps: 32

US-09-916-849a-3 (1-2923) x AK646739 (1-24370)

QY 1 MetatserProAlaThrGlyValProleuProThrProProProleuLeuLeu 20  
 DB 201 ATGCGGAGCCCGGACCGCGGCTCCCTCCCAACGCCGCGCGCTGTGTG 260  
 QY 21 LeuLeuLeuLeuLeuProProProleuLeuGlyValGlyProCyahrsSerleu 40  
 DB 261 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320  
 QY 41 GlySerArgGlyArgGlySerSerGlyValaCyahrsAlaProMetGlyTyrProleuCyahrsProser 60  
 DB 321 GGGTCACAGGAGCAAGGCTCTTCGGGGGCTGCGGCCCAATGAGGCTGTCTCAATCC 380  
 QY 61 SerAlaSerAmeLeuTyrTrpLeuTyrThrsSerArgCyahrsAlaGlyTyrGluLeuThr 80  
 DB 381 TCAGGCTGCAACTCTGCTCTACACAGCGCTGCAAGGAGTGGGGGCACTGAGTGA 440  
 QY 81 GlyHisLeuValProHisHisAlaArgGlyLeuArgValTyrCyahrsProGluSerGluAlaHis 100  
 DB 441 GAGCACCTGGTACCCACACAGATGCTGAGGGTGTGTGTGTGTGTGTGTGTGTGTGT 500  
 QY 101 IleProleuProProAlaProGluGlyCyahrsProTyrSerCyahrsLeuLeuGlyTyrGly 120  
 DB 501 ATTCCCTTACACACAGCTCTCTGAGAGCTGCTGAGCTGTCTGCTGCTGCTGCTGCT 560  
 QY 121 GlyHisLeuSerProGluGlyValLeuLeuThrLeuProGluGluHisProCyahrsLeuVala 140  
 DB 561 GGCCACCTTCTCCCAAGGAGCAAGCTCACATGCTGCGGAGAGCACCGTGTCTTAAAGCT 620

QY 141 ProArgLeuArgCyahrsLeuSerCyahrsLeuValaGluAlaProGlyLeuArgAlaGlyGlu 160  
 DB 621 CCAAGGCTCAAGATGCAAGTCTCTGCAAGTGGCAAGGCCCCCGGCTCAAGGCGAGGGA 680  
 QY 161 ArgSerProGluGluSerLeuGlyGlyValArgGlyValArgAlaValaThrAlaProGlu 180  
 DB 681 AGGTTCACAGAAAGTCTCTGAGTGGGCTGAGAAAGAAATGTAAATACAGCCCCCAG 740  
 QY 181 PheGluProProSerArgTyrGluAlaThrValProGluAlaGluProAlaGlyTyrProVal 200  
 DB 741 TTCAGCCCCCAGCTACAGGCAAGTCTGAGGAGAACAGACAGAGGCAAGGCAAGGCTGT 800  
 QY 201 AlaserLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluTyrThrMet 220  
 DB 801 GCATCCCTGAGAGGCAATGACCCCGGACGAGGAGTGAAGGAGTCACTGAGTGAAGCAACAT 860  
 QY 221 AspAlaLeuPheAspSerArgSerArgGluPhePheSerLeuAspProValThrGlyAla 240  
 DB 861 GATGCCCTCTTGTATGACCGCTCCACACAGTCTTCTCCAGAGCCAGTCACTGTGCA 920  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrIleSerThrHisValPheArgValThr 260  
 DB 921 GTACACAGCCGAGAGTCTGATCTGTGAGACAGAGACACCAAGCTTCAAGGCTGACG 980  
 QY 261 AlaGluAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 981 GCGCAGAGCACCGGCAATGCCCCGACGAAAGTCCCTGCTACCTCACTCACTCTTGATTACT 1040  
 QY 281 AspThrThrAspHisAspProValPheGluGluGluGluGlyValGlyGluSerLeuArgGlu 300  
 DB 1041 GACACCAATGACCAATGACCTGTGTGTGAGAGAGAGAGTCAAGAGAGCTCAGAGAG 1100  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 DB 1101 AACTGAGAGTGTGCTATGAGTCTCACTGTACAGGCAAGATGATGATGATGATGATGAT 1160  
 QY 321 AsnAlaAsnIleLeuTyrValGluLeuLeuGluGlySerGlySerProSerGluValPhe 340  
 DB 1161 AATCCCAATATTCGTGACCCGCTGCTGAGAGGCTGTGGGGCAGACCCCTCGAAGTCTTT 1220  
 QY 341 GluIleAspProArgSerArgValIleArgThrArgGlyProValaAspArgGluVala 360  
 DB 1221 GAGATCAACCTCTGCTGAGGATATCCGAAACCGTGTGCTGTGATCGGAAAGAGGTG 1280  
 QY 361 GluSerTyrGluLeuThrValGluAlaSerAspGluGlyArgAspProGlyProArgSer 380  
 DB 1281 GAATCTTACAGAGTCAAGTGAAGGCAAGTGAACAGAGGTGGAGCCGGGTCTCGAGAT 1340  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGluPheSer 400  
 DB 1341 ACCACACCCCTGTCTTCTTCTGTGTGAGAGATGACAAATATATGCCCCCAGATTATGT 1400  
 QY 401 GluIleAspArgTyrValaGluValaArgGluAspValaThrProGluValaProValLeuArg 420  
 DB 1401 GAGAGGCTATATGTGTCTCAGGTGAAGAGATGTACTCAGAGGCCCCAGATCTCCGA 1460  
 QY 421 ValThrAlaSerAspArgAspArgGlySerAsnAlaValValHisTyrSerIleMetSer 440  
 DB 1461 GTCAACAGCTCGATGAGCAAGAGGAGCAAGCCGTGTGTGTGTGTGTGTGTGTGTGTGT 1520  
 QY 441 GlyAsnAlaArgGlyGluPheTyrLeuAspAlaGluThrValaLeuAspValaValSer 460  
 DB 1521 GGCATATCTGGGGAGAGTGTATCTGTGAGTCCAGACTGAGAGTGTGTGTGTGTGTGTGT 1580  
 QY 461 ProleuAspTyrGluThrThrIleGluTyrThrLeuArgValaArgAlaGluAspGlyGly 480  
 DB 1581 CTTCTTATCTATGAGAGCAAGAGATGACCTTACGGGTGCAAGCAAGATGATGTGCTG 1640  
 QY 481 ArgProProleuSerAsnAlaSerGlyLeuValThrValaGluValleuAspIleAsnAsp 500  
 DB 1641 CGTCCCACTCTCTATATGTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1700  
 QY 501 AsnAlaProIlePheValSerThrProPheGluAlaThrValleuGluSerValProleu 520

```

Db      1701 AAGGCCCAATCTTCGACAGACCCCTTCAGGCTACTGCTCCGAGAGTGTCCCTTA 1760
Qy      521  GATYrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGluAspAsnAlaArgLeu 540
Db      1761 GGGTACTGTTCTTCATGTCCAGGCTATGACGCTGATCTGTGGTGAACAAGCCCGCTG 1820
Qy      541  GATYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db      1821 GAATACCGCTCTGCGGTGGGACATGACTTCCCTTCACCATCAACATGGCACAGGG 1880
Qy      561  ThrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      1881 TGAATCTCTGTGGCTGTGAATGAGACCGGAGAGAGTGAATTTCAACACTTGGGGTA 1940
Qy      581  GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerAlaSerValThrVal 600
Db      1941 GAAGCTGAGAACCATGGCACTCCAGCACTCACTGCTCCGSCCAAGTCAAGCTGACTGTC 2000
Qy      601  LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      2001 CTGAGATCAACCAACCAACCACTTACCAACCAAGTACACAGTGGCGCTCAAT 2060
Qy      621  GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      2061 GAGGATGACGCTGTGGCACACGAGTGTGACGCTGTGACGCTGTGACCTGTATGCTCAT 2120
Qy      641  SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db      2121 AGGTATCATCACTTACCAAGATCAACAGTGGCAATCTGMAACCGCTTCTCCATCACACAG 2180
Qy      661  GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGlnArgGln 680
Db      2181 CAAGAGTGTGTGGTGGCTGTATCTTGGCCCTGCCACTGACATCAAACTTGGAGCGGACG 2240
Qy      681  TyrValLeuAlaValThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleValVal 700
Db      2241 TATGTGTGCTGTTCACGCTCCGATGAGCACTGGAGAGACAGGACAGATGTGGTG 2300
Qy      701  AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      2301 AATGTACCGACCGCAACCAACCAATCGTCTTTCAGAGCTCCCACTTACAGTGAAT 2360
Qy      721  ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db      2361 GTTAAATGAGACCGGCGGACGACCAAGTGTGTATCAAGCCGACGAGTGAAGTAC 2420
Qy      741  ThrGlyGluAsnAlaArgGlyIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      2421 ACAAGTGAAGATGCCGACATCACTTCTCATGAGAGACAGCAATCCCACTTCCGATC 2480
Qy      761  AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      2481 GATGACAGACACGGGGGCTGTCAACCAAGCTGAGTGAATGAGACCAAGTGTCT 2540
Qy      781  TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800
Db      2541 TACACCTTGCCATTAATGTCTCGGACAAATGAGCAATCCCAAGATCCGACACACTTAC 2600
Qy      801  LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyr 820
Db      2601 CTGAGATCTCTGTGAACGACGTGAATGACAAATGCCCTCAAGTTCCTGACAGACTCTAC 2660
Qy      821  GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      2661 CAGGGCGTGTCTATGAGATGTGCCACTTCACTTGAAGTCTCTGACGATCTCAGACACT 2720
Qy      841  AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db      2721 GATGTGATTTCTGACTTAATGAGGGCTTCTTCAACCTTCCAGAGGCGACGATGGA 2780
Qy      861  AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880

```

---

```

Db      2781 GACGGTGACTTATTTGTTGAGTCCAGTCAAGCATCTGTGCGAACCTTACGAGCGTGAT 2840
Qy      881  ArgGluAsnValAlaGlyTyrValLeuArgAlaTyrAlaValAspLeuGlyMetProPro 900
Db      2841 CAGAGAAAGTGGCCCAATATGTCTTGGGCAATGAGATGACAGAGGAGATGGCCCCA 2900
Qy      901  AlaArgThrPrometGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db      2901 GCCCGCACACTTATGAAAGTGAAGTCACTGTGTGTGATGTGAATGACAAATCCCTGTTC 2960
Qy      921  PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      2961 TTTGAGCAGATGAGATTTGATGTGTGTGTGAGAGAGAACGCCCAATGGGCTTACCGCTG 3020
Qy      941  AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      3021 GCCCGGTCAACGACCTGACCCCAAGAACCAATGAGGACCAATGCCAATTTATGACAGAT 3080
Qy      961  ValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla 980
Db      3081 GTGAGAGGCAATCCCTGAGGCTTTCAGCTGACATCTTCTCCGGGAGCTGACAGGCC 3140
Qy      981  LeuValAspLeuAspTyrGluAspArgProGluTyrValIleValIleGlnAlaThrSer 1000
Db      3141 CTGTGAGACTTGAAGTACGAGGACCGGCTTGAATGATGATCTGTGATCAACGACCTCA 3200
Qy      1001  AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020
Db      3201 GCTCTCTGTGTGAGCGGCTTACACTCACTCACTCCGCTCTTGAACGCGCATATCAACCA 3260
Qy      1021  ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040
Db      3261 CCAAGTCTGGGCAACTTGAAGATCTTTTCAACCACTATGACCAATGCTCAAGACAGC 3320
Qy      1041  PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      3321 TTCCCTGGGGGTGCCATTTGGCCGATACCTGCTGCCAATGACCTGATATCTCAATATGCTG 3380
Qy      1061  ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      3381 ACTTACAGCTTGAAGCGGGGAAATGAATCAAGCTGTGCTCTCAATAGCTCCACAGGAT 3440
Qy      1081  GlnLeuValLeuSerArgAlaLeuAspAsnAsnAspProLeuGluAlaIleMetSerVal 1100
Db      3441 GAGCTGAAGCTAAGCGCGGCACTGACAAACACCGGCTCTGAGAGCCATCAATGACGTG 3500
Qy      1101  LeuValSer----- 1103
Db      3501 CTGTGTCT-AGGTAAAGAAAGGCCAGGTGGCGCTGGGGTGGGGGTAGCTCGCGGGAGAT 3559
Qy      1103  ----- 1103
Db      3560 GTCTGGGACGCACTGAGATGAGGTGCGATCCAGAAAGACGTAACATCCACTCCT 3619
Qy      1103  ----- 1103
Db      3620 GCCCAGTGCCTGGCAACAGACAGAGGGGCAAGACCTTGGAGAAAGGCCCAAGAA 3679
Qy      1103  ----- 1103
Db      3680 CCTGGCGGCTGTGTGACAGGCCCACTTCTGTGTGGGCTGTGGGCTGTGCTTCGTGGTGGC 3739
Qy      1103  ----- 1103
Db      3740 TGCCCCACTGTCTTACAGACCTGTGACGTGTCTGGGAAAGTGTGTGAGGCTGCCCGGC 3799
Qy      1103  ----- 1103
Db      3800 TATGGCTGCAGAAATTTGAAAAAAGCGCGTGAAGCTTGGGGGCGCAAGAGACTGG 3859
Qy      1103  ----- 1103
Db      3860 GGAAGGCGACAGAGACGTGTGCGCAGAGAGAGAGGCGAGGCGCAAGGCCCTTCAACCC 3919

```



Db	6080	GAGAAAGTGTTCCTCCAGGGCTGAGGGGATG6CAAAAGTTTCCTCCGCTCCATGSCCTTGG	6139
QY	1103	-----	1103
Db	6140	TTGCCCCCTCTCTCTCTTAAATCAGTTATGAGGCACTGTGTGTGAGATCTTGGGCTGCTT	6199
QY	1103	-----	1103
Db	6200	CTGCTTGAGGGAATATCCCGTCAAGAAATCATTAAGCAGATATACAGGTATACAAAG	6259
QY	1103	-----	1103
Db	6260	CTCTGTGCGACGTGAGCATCTGTGTGCGGGAACTTGAAACGAGAGAAAGTAGGCC	6319
QY	1103	-----	1103
Db	6320	TGATGACATAAATCTGGCAGGCTTCTTGAGCAGGTGAGCTAAGGCGAGTTAGAAGCT	6379
QY	1103	-----	1103
Db	6380	CAGAGACTGTGTGGGTGTGTCTGGGTGGAAGAAACGCTGTGTGACAGCTTTGAGAC	6439
QY	1103	-----	1103
Db	6440	GAGAGCAGAGAAACAGATGTGGTGATGACAAGCAGGCTTGCTGTGTGAGACTAGTGAG	6499
QY	1103	-----	1103
Db	6500	AATTAAGCAATGATGAGAGGCTGGAAGGTATGGTGGCAGTTCGGCTAGGACATGGCTGAGAT	6559
QY	1103	-----	1103
Db	6560	GAGGGTCTGCTTGACCTTCCGGCTTCAGCTGAGCACCGCTGTATCAGTACAGCAC	6619
QY	1103	-----	1103
Db	6620	GCAAGTAGATACAAACACAGATGACGAATTTGTTCCATGTCCCTTGCTACTCCACCC	6679
QY	1103	-----	1103
Db	6680	ACCCAGAAATGAGGCTCTGCCCCGTACTGCTAAGGTAAGCATTTGCTTTTAAATC	6739
QY	1103	-----	1103
Db	6740	TAGTTGGGCTCCCTGTATGTGTGTTGTTGTTGTGTTGGACAGAGTCTCACTCG	6799
QY	1103	-----	1103
Db	6800	TCACCCAGGCTGAGATCAATATGGCGCATCTCGGCTCACTGCAACTCTGCTCTGGGT	6859
QY	1103	-----	1103
Db	6860	TCAAGCAATCTCTGCTCAGCCTCTGAGATCTGGAGACTCTGAGTAGTAGAATT	6919
QY	1103	-----	1103
Db	6920	ACAGGACCCACCAACAGCCCACTAATTTTGTATTTTATAGTAGAGACGGGTTTCAC	6979
QY	1103	-----	1103
Db	6980	CATGTGTGATCAGTCTGATCTGAACTCGTGAACCTGTGATCAACCGGCTCAGGCTCCCA	7039
QY	1103	-----	1103
Db	7040	AAGTTCTGGAGCTACAGCGGTGAGCCACTGGCGCCGCTGTATGTTTAAAGTCC	7099
QY	1103	-----	1103
Db	7100	GTTTATTTCTTTGTGAGACTGAATAGATTTGATTCATCAGATCTCAGTGTTTCTGAACAC	7159
QY	1103	-----	1103

Db	7160	TACTATGTGAAGGCATTGTAAATAAATCAAAATATGAAATATGCAGTAGCTCATTCCTGT	1103
QY	1103	-----	1103
Db	7220	AATCCAGTGTCTTAGAAGCCTGAGATGAGAGATCACTTGAGGCCAGAAATTGAGACCA	7279
QY	1103	-----	1103
Db	7280	GCTTGACACACAGCAAGAACCCCATCTTACACACAAACAAAATTAGCCAGGATAGT	7339
QY	1103	-----	1103
Db	7340	TGTGAGCAGCTGTGTCTCAGCTATTTGGAGGTGAGGTGGAGAGATCGCTTGAGCCA	7399
QY	1103	-----	1103
Db	7400	GGAATTGAGGCTACAGTAGCTATGATCGTGCCACTGCTCTCCAGTGTAGGCGACAGAG	7459
QY	1103	-----	1103
Db	7460	AGTCTGTCTTTAAAAAAGACATATCCCTCTTGCCCTGTAAAGCCATTT	7519
QY	1103	-----	1103
Db	7520	CCTCCTTCTCCAGGCCCCAGCTTTCATTTTGTAGTAAATAAATTCAATCCATGTGTA	7579
QY	1103	-----	1103
Db	7580	TGTGCATTCAAGGTGTATGGTATTTCTGGGCTTAGGGTGTGTGGGTGGAATGAGACCC	7639
QY	1103	-----	1103
Db	7640	AGCAAGTGTTTTAGTCTTAGCTGTATTAAATGATGATGTGTCTCGGGGCGATAGCA	7699
QY	1103	-----	1103
Db	7700	GGCATTGGGTTATATGGTTGAGCCAAAGTGAGTATGATAGTAGAGCCCTTAGACCA	7759
QY	1103	-----	1103
Db	7760	GAGAAAGACTGACTCAGTGGGGCTTGGGGTCTCAGAGTGAAGAGCTGCCAGCCCTT	7819
QY	1103	-----	1103
Db	7820	CTATCTCTGCTCCGGTCCCCACCCCACTCACTGTCTCCAGCATGGGGCAGTGGG	7879
QY	1103	-----	1103
Db	7880	TGATGATGTCACTGTGTTAAGTGGCTCAGCTGATTTCTGGGGTGGGGCCCCCAAGCT	7939
QY	1103	-----	1103
Db	7940	ACCACAGACAGCCAGAGAGTTGGGCGGGAAGATGTGTGTGATGTTCATGGG	7999
QY	1103	-----	1103
Db	8000	GATAGCAGGTCTGGCGCCCAAGACGCAAGGGCTCAGTGGAGGGGGACTAGTGGCC	8059
QY	1103	-----	1103
Db	8060	CACTAGCCAGTTTACCACTCCCCCACCAGCACTCCCTGTGCTGTGGGGAG	8119
QY	1103	-----	1103
Db	8120	TGGGGAACACAGCCATGGTTGCCATTATTCGTGTGTTGAGGTGTAGAGGCTGAGGGGCA	8179
QY	1103	-----	1103
Db	8180	GGGGGATGATTTTGTGGGTCTCAGATGTCTGTGGCCCTGTAGAAAATGTGAAAAAGA	8239
QY	1103	-----	1103
Db	8240	CAACATAGGGTTCCCGGCGAGGACGACAGCTCTGGGGGAAAGGAGCAGAGCGGTGCG	8299

QY	1103	-----	1103	QY	1319	-----	1319
Db	8300	TTGGCAGGGAAGGTGCACAAATTAGAGGTGTCCGTGCATTCAGCAGACAAAGTT	8359	Db	9380	GGAGTGTCTTCCCTGCTGTGCACAGACACAGAGGAGTGCATTGTGGCAGAGCTGG	9439
QY	1103	-----	1103	QY	1319	-----	1319
Db	8360	TTGGGAAGAGATGCAGAAGAACTGCMAATCTCATCCCTGAGCCCTGAAAGAGTGAATTG	8419	Db	9440	CTCCAGACCAATGCGCCTGTGTTCGTGTATCAGCTTTGCCGTAGTAGTGTGACCT	9499
QY	1103	-----	1103	QY	1319	-----	1319
Db	8420	GCCCAACGTGACATPACTGTGGCTCAGACCCCTTATGCCAGTTGAAAGAGACCTTTCT	8479	Db	9500	TGAACAAGTTACTTAACCTTTCTGGCCCTCGGTTATTGAAATGAGATTAATGACCTGAC	9559
QY	1103	-----	1103	QY	1319	-----	1319
Db	8480	TGGTGGTGTGTGCAGGAATATGCAGGCTGAATTGAGCCCGAGGCTGGTCTGACTGTG	8539	Db	9560	TTAGGTCAATAAGTTGGTTGTGAGGATTAATGTAATTATTTGATACCTTGAAACAGCTGG	9619
QY	1104	-----	1119	QY	1319	-----	1319
Db	8540	TTCCCTCTCCACAGACGCGCTGACAGCCGTGACCCGCTGAGGACATGTACACCCGAG	8599	Db	9620	TGGCATGGAAGTGTGGCTTGATTTAAGGAAAGCTGAGATTTGGTGACTTCTCTCTG	9679
QY	1120	-----	1139	QY	1319	-----	1319
Db	8600	ATCAGCGATGAGATGCTCACCCACAGCATCACGCTGCGCTGAGAGCATGTACACCCGAG	8659	Db	9680	GTTCTATCACAACCCGCTTTAGAGGTGAGGCTGGCCTAGAACTGTGTCTCCCCCACTG	9739
QY	1140	-----	1159	QY	1319	-----	1319
Db	8660	CGCTTCTGTGCACCACTGTAGGCTCTTCATCCAGGGGGTGGCCGCCACGCTGGCCACG	8719	Db	9740	AGCGATTCCTTTCTGAGCTGACAGAAATCTTTCTCCCAAGCATTAAGGGGAAGAC	9799
QY	1160	-----	1179	QY	1319	-----	1319
Db	8720	CCACCGGACCAAGTGTGTCTTCAACGTACAGCCGACACCGACGCCGCCGGGGCCAC	8779	Db	9800	ACAGAAAGCTGGGGAACTGAGGGGGTTGAGTTCTGAGATGGGTGCTCCAGAAAAAC	9859
QY	1180	-----	1199	QY	1319	-----	1319
Db	8780	ATCTCAACGTAGAGCTGTGCGGTGGCCAGCCGACGGGCCGGGGGGCGGCCCTTC	8839	Db	9860	ATTGCACTGTATGGGCTGTGGCAGAAATGGCTCCGCTCCGCCAGCTGCTGTAATAAGA	9919
QY	1200	-----	1219	QY	1319	-----	1319
Db	8840	CTGCCCTCTGAGGACCTGCAGAGAGCGCTTATCTCAACCGACGCTGTGACGCGCATTC	8899	Db	9920	GGATCTGTGGTAGGCGGCTTCCCTGTGGCTGTGACGCAAGTCTTTACATCTGTCT	9979
QY	1220	-----	1239	QY	1319	-----	1319
Db	8900	TCGGCAACGCGGCTGTGCTTTCGACGACATCTGTGCTGCGGAGGCCCTGCCAGAAC	8959	Db	9980	CCCTGGATGTCACTGTCTGTCTTATGTCTGTACCCCTCCCTTCAAGCTCTGTGT	10039
QY	1240	-----	1259	QY	1319	-----	1319
Db	8960	TACATGCGCTGTGCTGTGCTGCTTGCATCTCTCCGCCCTTCTATGCGCTCTCTCC	9019	Db	10040	CTTCTCAGAGACACCCCTGCTCCCTTCCCTGTGGGCTTTCTTCTTATGATGCCAGGC	10099
QY	1260	-----	1279	QY	1319	-----	1319
Db	9020	TCCGTGCTCTTCCGCGCCATCCACCCCGTGGAGGGCTGCGCTGCGGCTGCCGCCGCC	9079	Db	10100	CCTGCTCTTGTCTCTCAGAGGCAATGCCCACTGGCTGTGCCAGCAGCATATGAGGAGGCC	10159
QY	1280	-----	1299	QY	1319	-----	1319
Db	9080	TTCAAGGCTGATCTACTGCGACGCGAGGTGACCTGTCTACTCCGCGGCTGTGTGCCCC	9139	Db	10160	CATCTCTTCCCGCTGATCTTCTGTCCAGAACTTGCCAGACTTGACTTCTCTGTCTG	10219
QY	1300	-----	1319	QY	1319	-----	1319
Db	9140	CACGGGCGCTGCGCAGCGCCGACGGCGGCTACCTGCTTGTGTGTATGAGCTACACG	9199	Db	10220	CTGCTAAGGGTGGGGGAAGGGAGGAGCGGGCTTGGGGGAGGGGAAGAAAGTCTTTGTC	10279
QY	1319	-----	1319	QY	1319	-----	1319
Db	9200	GGTGAGCCAAAGGAGGAGCTCATGGGCGAGCCCTGAAAGCTGATGTGTGTGAGGC	9259	Db	10280	CAGAGGATGTGTGGGGGGTGTGGAGGGGGGGGTTAGAGCGAGCGAACAAGATTTC	10339
QY	1319	-----	1319	QY	1319	-----	1319
Db	9260	ACAAATCAGGACAATGTGCGGGCTGCTCATTTCTTCCGAGTGAAGTGAAGTGA	9319	Db	10340	ATGAAACAAAGAGCCAGTCACTAGGCAAGAGAGTGAACCTATCTTGAATCTGGGA	10399
QY	1319	-----	1319	QY	1319	-----	1319
Db	9320	CTGAGAGGTCAATAATCCGGCTGTCTCTGACAGACCCCACTGAGAGAGCAATGCA	9379	Db	10400	ATGGGAGGAGAGAAAGGTGGAATTTTCCATCTGATGTTGAGTTTTCAGCTCTCTGT	10459

Db 10460 TGAATTTCTGTCTCTGCTCTGAGGCTCAAGTTTCCACTCTCAGCATGTCCAGGTGCACAG 10519  
 QY 1319 ----- 1319  
 Db 10520 AGGTCTTAGTCTGACCGCAGGGCTCAAGGTGGAGTCTCTGCCCCATTCTGGGGGC 10579  
 QY 1319 ----- 1319  
 Db 10580 TCCCGACTTCTCTCATACTCTTCCATTAACCAAGAACTAAGTGAAGCTGTGC 10639  
 QY 1319 ----- 1319  
 Db 10640 CCATCCCGAGGAGGAGAACTGATCTCAGTGCCTGAGAGGACAAAGACCAAGTGGAA 10699  
 QY 1319 ----- 1319  
 Db 10700 GGAAGGAGGAGGAGGCTCCAGCCCTCCCGCTGTCTGCAAGGCCATCTGCTGTGCAC 10759  
 QY 1319 ----- 1319  
 Db 10760 ATTTAACCTCTTGGGGGCGAGGACACCCAGGTGTGTTGGTAGTGAATCGAAGGC 10819  
 QY 1319 ----- 1319  
 Db 10820 TCCCGGATCCGGGCTTAAACAAGTGTCCAGAGGCTGTGCTTAATGAACCCAA 10879  
 QY 1319 ----- 1319  
 Db 10880 TGGGGTAGGGAAGGTCTGGGACTCCAGAACTTGCAGGGGAGCAAGTGTGAGCTG 10939  
 QY 1319 ----- 1319  
 Db 10940 TGGCAGAGCGGCTCAGTGAACCAAGGTGGCCATGAAGGCTGGGACCATAGGTC 10999  
 QY 1319 ----- 1319  
 Db 11000 TCAGAGAGATGACACAGAGAGAAATGGGATGCTGGAGGTTTGGGAGTATTACG 11059  
 QY 1319 ----- 1319  
 Db 11060 AGGTCAAGGAGGCGCATGGAGCTATCATGAGGTTTGGGGGCTCAGTTCCCTCTCT 11119  
 QY 1320 ----- GlyGluHisCysGluVal 1325  
 Db 11120 GTCTCAGGTACTGTGACTCACTGCTCTTCTCTTGTCCAGGTGAGCATGTGAGGTC 11179  
 QY 1326 SerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGlyVal 1345  
 Db 11180 AGTGTCTGCTCAGGCGGTGACCCCGGAGTGTGAGAGATGGGGGCACTGTGTCAAC 11239  
 QY 1346 LeuLeuValGlyIlePheIleCysAspCysProSerGlyAspPheGlyIleValProIleCys 1365  
 Db 11240 CTGCTGTGTGGGGGGTTCAGAGTGCATTCCTCATCTGAGAGCTTGGAGAGCCCTATGTC 11299  
 QY 1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyIleLeuGln 1385  
 Db 11300 CAGGTGACCAACCGCAGCTTCCCGGCCACTCTTCATCACTTCGCGGCTGTGGCCAG 11359  
 QY 1386 ArgPheHisPheThrLeuAlaLeuSer----- 1394  
 Db 11360 CGTTTCCACTTCACTGAGCCCTCTTC-GTGAGTGGCTGGGGCACTGGGGGTGGGAGTGGG 11418  
 QY 1394 ----- 1394  
 Db 11419 CCTGTGGGCACTAGAGTGTCTCTGTCTGTGTGTTAAGTAAGAGTCAAGCAGTGA 11478  
 QY 1394 ----- 1394  
 Db 11479 AGCGTGAAGCAGGCTGCCGAGAGCCAGGTCTGCTCTTGAAGTTCCTGTGGTTCTGCG 11538  
 QY 1394 ----- 1394

Db 11539 AGCAGATCTCCCTCCCTGCGTGGTTAAGTGGCGGGGTGCTGCATTCAGGTGGTAC 11598  
 QY 1395 ----- PheAlaThrIleGlyValArgAspGly 1402  
 Db 11599 CCATTCCTGCCCCCATCCCAACTCTCTGTTCAGGTTTGCACAAAGAGCGGAGCGG 11658  
 QY 1403 LeuLeuValIleValArgPheAsnGlyValHisAspPheValAlaLeuGluValIle 1422  
 Db 11659 TTGCTGTGTACATATGGGCGTTTCATATGAGAACTGATGATGTCCTGAGGTGATC 11718  
 QY 1423 GlnGluValIleGlnLeuThrPheSerAla----- 1432  
 Db 11719 CAGAGCAGTTCAGCTCACTTCTCTGC-AGGTATCAGATGTCCTCCCATCTTGC 11777  
 QY 1432 ----- 1432  
 Db 11778 CATCTCCAAAGGCCCAAGTCTTCAGCCCTGACCCCAAGCCATCTATCAGCC 11837  
 QY 1432 ----- 1432  
 Db 11838 AAATCTGGGCCAGCCCAAGCCCACTGGACCCCAACCTGCAATCTAGCCTGTCA 11897  
 QY 1433 ----- GlyIleSerThrThrValIleSerProPhe 1442  
 Db 11898 GGCATTCAGCTCACTGTCTCTTCCCGAGGAGTCAACACAGGTGTCCCATTC 11957  
 QY 1443 ValProGlyIleValSerAspGlyIleThrPheThrValGlnLeuValIleVal 1462  
 Db 11958 GTGCCGAGAGTCAATGATGGCAGTGGCATTCGCTGACGTGAAATCTAATTAAG 12017  
 QY 1462 ----- 1462  
 Db 12018 GTGGGTGAGAGGCAAGAGGTTGGGGTCTGTCTTGTCTCAGGTGCTTACCAAGC 12077  
 QY 1462 ----- 1462  
 Db 12078 CTTAGTGGCATTTGCTTCAGGCTTGGGTGGCTGTCAAGGCAATTTGCTGAGAGA 12137  
 QY 1462 ----- 1462  
 Db 12138 AGGAGGCGGTGGGAAGTGCACCTTGTGGGATTCACCAACCCACATGGGCACTGCTG 12197  
 QY 1462 ----- 1462  
 Db 12198 ATGCTCTGTTCATAGGACCGCAAACTTGAACATGAGAGATGATTAATGCA 12257  
 QY 1462 ----- 1462  
 Db 12258 TAAACCATGAGCAACGTGTGTATATGATGATGTGATGATGTGTGAGGATGTG 12317  
 QY 1462 ----- 1462  
 Db 12318 GTGACACAGAGGACCCAGATGTGTGCTGGCATTAGTGGGCTGTGACTAGTG 12377  
 QY 1463 ----- ProLeuLeuGlyGlnThrGlyLeuProGln 1472  
 Db 12378 TCCCTCTCTGCTCTTCTCTGTCACAGGCATGTTGGGTGAGAGCGGCTCCACAG 12437  
 QY 1473 GlnProSerGluGlnValAlaValIleValIleValIleValIleValIleValIle 1492  
 Db 12438 GGGCCATCAGAGCAAGAGTGGCTGTGTGATCCTGTGATGCTGTGATGATGATGATG 12497  
 QY 1493 LeuArgPheGlySerValIleLeuIleAsnIleSerCysAlaIleGlnIleGlnIleGly 1512  
 Db 12498 TTGGCTTCGATCTGTCTGTGGCACTACTCTGTCTGTGCGGACCCAGGAGTGGC 12557  
 QY 1513 SerIleValIle----- 1515  
 Db 12558 AGCAAGAGTGAAGCAGGAGAAAGGCCAGGATGGGTGAAGTGAAGGCGAGGCCAGGCA 12617  
 QY 1515 ----- 1515  
 Db 12618 GGAAGGAGGTGAAGAGGGGCTTTTGGGATCTGTCTCTCTTGTGCTCTTACGCC 12677



QY 1515 ----- 1515  
 Db 12678 CTGCTCAGCCCTGGGGATGAGGATCAAGCTGAGGAAATCAGAGCAAGCCCTGTGGGGCAG 12737  
 QY 1515 ----- 1515  
 Db 12738 TCTGGGGAGGGGGCTGAGTGTTCAGCAAGAAACAGTTCCAGGTAGGGGTGACGCCCTTCTCT 12797  
 QY 1515 ----- 1515  
 Db 12798 GAGCACTGCTCTGAGAGGGCTCTTGGAACCTGCTACGCGGTTCTGTCTCTGAGCCCTG 12857  
 QY 1515 ----- 1515  
 Db 12858 CGGCAAGCGAGCGAGGACCAAGAGCTCCAGCCCTCTGTGCGGCCACCACTCTTG 12917  
 QY 1516 ----- 1516  
 Db 12918 CTCCCCCAGGCTCTCTGAACTGACCGGGCCCTGCTACTAGGCGGGGTGCTGACCTG 12977  
 QY 1532 OGUSerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAs 1552  
 Db 12978 CGAAGAGCTTCCAGTCCGAATGCGGCAAGTTGCTGGGCTGCAATGGGAACCTGCAAGTGA 13037  
 QY 1552 pSerTrgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrVal----- 1568  
 Db 13038 CAGCCGGCACTAGACATGCTGACTTCATTGCAACATGCAACCTGCTGTATGAG 13097  
 QY 1568 ----- 1568  
 Db 13098 GGGCCGGGGTGGAGCCAGGCTGGGATCCCACTGCTAGAGAGAGGGGCTGTGCTGTG 13157  
 QY 1568 ----- 1568  
 Db 13158 AGATGCGCAGAAATGAGAGGTCTCAGAGGCTCTCTAATGCTGCGACAGGCTTCTCTTT 13217  
 QY 1569 ----- 1569  
 Db 13218 AGAGCCCAAGCTCAGGTCCACTGAGCTGCTTCTCTCAATCCAGGCTGCGCTCTCCAGAGA 13277  
 QY 1575 YAsnValCysAspSerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAsp 1595  
 Db 13278 AGAACGTGTGTGACAGCAACCTTGCCACATGAGGGGCACTTGCTGTGACAGTGGAGCG 13337  
 QY 1595 IAspSerCysGluCysProLeuGlyPheGlyGlyCysSerCysAlaGln----- 1611  
 Db 13338 GGTTCAGCTGCGAGTGCCTCCCTGGGCTTTGGGGCAAGAGCTGCGCCAGGGTATGAGAGG 13397  
 QY 1611 ----- 1611  
 Db 13398 GCGGCTGTAGAGGCAACAGCCTGGGTGCTCATCAACAGTCTGGAGACTGGCAGGGTGG 13457  
 QY 1611 ----- 1611  
 Db 13458 GCAGGCACTGGGCAAGGCTCTGTAGAGGGGGCTGTGGGTGAAAAGCGTGTCTGGCAG 13517  
 QY 1611 ----- 1611  
 Db 13518 AGCCCTGAAAAGGCAACACAGAGAGCAAAAGGGGCAAGTCAAGAGGCAAGGCTGGCACAAGC 13577  
 QY 1611 ----- 1611  
 Db 13578 GTGGGGGGGTCCAGGCAAGGTACGCACTTTGAGAGGCGGGGCTGATGAGGGAGTGGGC 13637  
 QY 1611 ----- 1611  
 Db 13638 TCTGCTCCCGTGCAGCGCCACCGCTGAGCATCAAGCCCAAGGCGCTCTGAGGCTCC 13697  
 QY 1612 ----- 1612  
 Db 13698 ACCGTCACAGTCTGCTTTCTGCTCTCCAGAAAATGGCCAAATTCACAGCACTTCTCTG 13757

QY 1621 GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrLeu 1640  
 Db 13758 GGCAGAGCCTGTGGGTGGCTGGCAATGAGCTCTCCCTGCCATCTCCAAACCTGTGATCCTC 13817  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 Db 13818 AGCCTCATGTTCCGACAGCCGACGAGCGAGGTGCTGTGAGGCAATCACAGGGGG 13877  
 QY 1661 ArgSerThrIleThrLeu----- 1666  
 Db 13878 CGCAGCACTATCAACCTCTACAGGTATGATGAAAGGGCGGCTGGCCCTGGCCATA 13937  
 QY 1666 ----- 1666  
 Db 13938 GGGCCCTGTAGCCTTAGCGGCTGAGCAGAAATGGCTGGGCAAGTCTGGCAGGGGT 13997  
 QY 1666 ----- 1666  
 Db 13998 GGGGACATATAGAGCGCGCTGATCCGTTGGGAAGTCAATGCTGCCACCTGTGGGCT 14057  
 QY 1666 ----- 1666  
 Db 14058 GAGGAAATATATACGCTTCTCTCTATGCGCTCAGCTGGGCAAGGGATGGGTTGAT 14117  
 QY 1666 ----- 1666  
 Db 14118 GGTCCAGAAAGGGAACTGTGAGCTGAGAACAGTATTAAGCGCTGTGACCTGT 14177  
 QY 1667 ----- 1667  
 Db 14178 TTTCTTTCCTCTCGGTGTCAAGTACAGAGGGCCAGGTATGCTGAGCGTGGAGGGCA 14237  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTrpHis 1700  
 Db 14238 GGGCTCAGGCGCTCTCTCGCTCGGTCTGAGGCAAGCGCGGCCAAATACCGTGTGCTGGC 14297  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 Db 14298 CATGCACAGCTGCACCTGGAGGCAAGCGGGGGCCCGGCATGCCATTCGTCTTCGAT 14357  
 QY 1721 TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 Db 14358 TATGGGCAAGAGACAGAGGCAACCTGGGCCCCCGCTGATGTCTGCACCTGAGC 14417  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyPheArgGly 1760  
 Db 14418 AACATTAAGTGGCGGAATACCTGGGCCAGCGCGGTGTGCCCCGTGGCTTCCGGGC 14477  
 QY 1761 Cys----- 1761  
 Db 14478 TGTTCAGAGTGAAGTGTCTGCTGCCCTGCCATCCCTCCCAACCACTGACAGCCCC 14537  
 QY 1762 ----- 1762  
 Db 14538 GCTCCACTAGAGGCAACACTGCTCTGTCTCTCCACAGGGGTGCGGTGAGCAATACGCC 14597  
 QY 1771 OGlnGlyValAsnSerLeuAspProSerHisGlyGlnSerIleAsnValGlnGlnGlyCys 1791  
 Db 14598 GGAAGGGGTTTAAACAGCTGGAATCCAGCATGGGAGAGACATCAAGTGAAGAAAGCGT 14657  
 QY 1791 AserLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAs 1811  
 Db 14658 TACCTGCTCTGACCTTGTGATCAAAACCGGTCTCTGTAAACGATATTCAGCAACGA 14717  
 QY 1811 pTrpAspSerTyrSerCysSerCysAsp----- 1820  
 Db 14718 CTGGGACAGCTATTCCTGACGCTGTGA--TCAGAGTATGCTAAGATTCAGGGCAAGGGC 14776  
 QY 1820 ----- 1820  
 Db 14777 AGGTATCAGTGTGCTGGGCAACATGCTGTGCTGTGTGTGACCTGGGCAATGGGCA 14836  
 QY 1820 ----- 1820

[illegible]

Db	15917	GCTCCTTGTCCAGAGTCGTGACCCCTGAGAGTGGCCAGTGTCCATGCAAGCCAGGTGCA	15976
QY	1950	legIaargInCyaaPaargCyaaPaasnProPhealIagIuaIThrThraengIyG	1970
Db	15977	TCGGGCGTCAGTGTGACCGCTGTGACAAACCCTTTTGTAGGTGACCAACAATGGCTGTGG	16036
QY	1970	lu-----	1970
Db	16037	AAGGTGGGGCTCCTGGGATGGTGGCAGCCCTCTTACATGTGTCTAGGCACCTGCACC	16096
QY	1970	-----	1970
Db	16097	CCAGAAAGCCAGGAAGGGGCTGTTCAGAGGCTGGGGAGGCTGGGAGGCTTAGGAG	16156
QY	1970	-----	1970
Db	16157	GGCTGGGAGCCTGGGAGAGCAGAGAGGAGAGCTGCTCTGGGTGACATGTGCTCT	16216
QY	1971	-----ValAasnYrAaspSerCyaaProArgalIaIegIuaIagIyleTPTpPro	1988
Db	16217	TCGCCGAGATGATTAATGACAGCTGCCCAAGACATTAAGGCTGGGACTGTGGCCCC	16276
QY	1988	rgIhArqPhegIyleuProalIaIaIaProCyaaProIygsIySerPhe-----	2004
Db	16277	GTACCCGCTTCGGGAGCTGCTGTGCTGCCCTGTCCAAAGCTCTTGTGTAGGTGT	16336
QY	2004	-----	2004
Db	16337	GGAGGCCCCCATGATGATGTGAGAGCATGAGCTCTGCTGTTAGTGGGATGAAGGACGAA	16396
QY	2004	-----	2004
Db	16397	GCTCCTGGCTGAAGATCCGGGAGACCCCACTTCTGGCTTCAGGTCTTGAGATTCAATT	16456
QY	2004	-----	2004
Db	16457	CCCATCTTCCTGGGTCAATCCCTTCCTCCAGATGTGCCAGGTGCTTATCTTTG	16516
QY	2004	-----	2004
Db	16517	GACTGTTTGTGTTTGTGTTTGTGTTTTCGAGGACAGAGCTGTGACTGTACC	16576
QY	2004	-----	2004
Db	16577	CAGGCTGTGNGAGATGGGCGTGAATCCGCTCGCTGCAACTCTGCTCCAGGTTCAAG	16636
QY	2004	-----	2004
Db	16637	CAATTTTCTGCTTAAGCCTCTGGCCGCCACACACTGGCTAATTTTTCCTTTT	16696
QY	2004	-----	2004
Db	16697	TTTTTTTTGTATTTTAGTAGAGACAGGGTTCAACATGTTGGCCAGTGTGTGAA	16756
QY	2004	-----	2004
Db	16757	CTCTGAACTCATGATCCGCTGCTCGGCTCCAAAGTGTGGATTACAGGCGTAG	16816
QY	2004	-----	2004
Db	16817	CCACTGACCCCAACCAATTTTGGACTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT	16876
QY	2004	-----	2004
Db	16877	TGTATCAGGCTACAAATGATGATGATCTTGCTCTTCGACCTTCACCTCAAG	16936
QY	2004	-----	2004
Db	16937	TTCAAGGATTTCTCTACTCACTCAACCTTCGAGTACGTGGTACTACAGGCGTGTCTACA	16996
QY	2004	-----	2004
Db	16997	TGCCAGCTAATTTTATATTTTTCCTTTTAGTAGAGACAGGGTTCAACATGTTGGCCAGGA	17056

QY 2004 ----- 2004  
 Db 17057 TGGTCGATCTCTTGACCTTGATCAAGCCGCTCGGCTCCCAAGTCCGAGATTA 17116  
 QY 2004 ----- 2004  
 Db 17117 CAGGCGTGAAGCAAGCGTCCAGCCACATCTTGACTTTAAGAAAAATGCTGA AAAAC 17176  
 QY 2004 ----- 2004  
 Db 17177 CTGCACTGTGCTTGCAATAGGCGTGTGTAGAGATTCAAAAGATGATGACAGGTTCC 17236  
 QY 2004 ----- 2004  
 Db 17237 AATCGTCACTCAACCAAGTCTGTAGATTAGCCACCAAAAGGAGACAGTCTGGG 17296  
 QY 2004 ----- 2004  
 Db 17297 AGTAGGGGGTGTGGCTGTACCCAGACACCCAGCCAGATGGGGGCTGGCGTGGT 17356  
 QY 2004 ----- 2004  
 Db 17357 CCTGACTGCAATAGACAGGTTCAATAGGAGAGAGTGGGGAGAGACAAAGTGTGGC 17416  
 QY 2004 ----- 2004  
 Db 17417 TGGAGACAGGAGAGATGGGGGCTAAGAGAGTGTATGCTGTCTCCCTAGAGCCATG 17476  
 QY 2004 ----- 2004  
 Db 17477 TGTGACATGATCTTGGGAGATTTTGGCCCTGGTGTCTAATACCTGTCTTCTGTTT 17536  
 QY 2004 ----- 2004  
 Db 17537 CACAGAGGCTTGTGTTGCTAGTTACTGTCCCGGCGCATGAGGTGTGGCTCCAGCAG 17596  
 QY 2004 ----- 2004  
 Db 17597 AACCTCTCAGCAAGTCTCAGGGGCTTCTGTGTCTCCCTGAGGCGGCTTTTGCT 17656  
 QY 2005 ----- 2005  
 Db 17657 TCCCTCCCGCAGGAGCTGTGTGCGCCACTGTGTAGAGACAGGGGGTGGCTCCCGCAA 17716  
 QY 2021 ~~g~~lyThraValAlaArgHisCysAspGluHisArgGlyTyrLeuProPro 2021  
 Db 17717 ACCTCTTCACTGACAGTCCATCACTTCTCAGAACTGAAGGGCTTCGTAAGTGAACCC 17776  
 QY 2036 ----- 2036  
 Db 17777 CTGATCTCATCTTTTCCCTGTCTTCGTCGTCGAGTCTGACTTGGCCCACTCCATCTT 17836  
 QY 2036 ----- 2036  
 Db 17837 TGAAGACGGGGCTTCTGAATTCAAGCTGTGTGTCTTGGGCTCCCACTGAGAGGCC 17896  
 QY 2036 ----- 2036  
 Db 17897 GTCTCTACCTCTGAGGTCCTTTTGCTCCAGAGTTCCGCTCATCTGCTCTCTCCG 17956  
 QY 2037 ----- 2037  
 Db 17957 CTCCTCTGTCTAGGCTGAGGGCTACAGCGAAATGAGTCAAGGCTTGAACCTGAGGGGC 18016  
 QY 2052 SerGlnGlnLeuAlaLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGly 2071  
 Db 18017 TCCGAGAGCTAGCCCTGCTCTGCGCAAGCCGACGAGACACAGCTGGCTACTTCGGC 18076  
 QY 2072 SerAspValIleValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGln 2091  
 Db 18077 AGCGAGCTCAAGGTGGCTTACAGAGCTGCGCAGCGGCTGTGGCCACAGAGACACCCAG 18136

QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGln----- 2106  
 Db 18137 CGGGCTTTGGGCTGTGTGCCACACAGAGCGTCACTTCACTGAGGTGGGCTTGAGAGA 18196  
 QY 2106 ----- 2106  
 Db 18197 TGCAGGGCTGGCTGTGTAGATAGGGGTCAATGTGTAGTAGCCTGTCTATGGCACTGGG 18256  
 QY 2106 ----- 2106  
 Db 18257 GGGCAGAGGGGGCTCTCCATCCACCTACAGAGAGCTCCCTGTGCGCAGAGGCTCCA 18316  
 QY 2106 ----- 2106  
 Db 18317 TGAACCTGTGACCCCTGGGCCAGCCCTTCCCAACCTTCTCATAGTAATATGTG 18376  
 QY 2106 ----- 2106  
 Db 18377 TGTGGGGCATGTGTGTGACCCAGACAGAGAGGCTGCTGTGTATCATGGCTGGAGAGA 18436  
 QY 2106 ----- 2106  
 Db 18437 AGCACTATCTAGAGCATGTGTGTGTGTCCAGGACAGAGGCTGGAGACTTATGTAGA 18496  
 QY 2106 ----- 2106  
 Db 18497 GAATGAGAGAGAGGCTTAGGGGAGAGACATCACTGACTGATGGGAGTGGCTCA 18556  
 QY 2106 ----- 2106  
 Db 18557 GGAGCATTTACAGGGAGGCAACAGTGAACAGAGGAGGGGCTGTGGGGAGAGAGGA 18616  
 QY 2106 ----- 2106  
 Db 18617 CTGGCCGGGGACAGGGCGGGGCTCCATGTGTGGGAGATGTCAAGTGTGGGTTGTAGACA 18676  
 QY 2107 ----- 2107  
 Db 18677 CCCACCGTGAACCTTGCCCAACCCAGAAATCTGCTGGGTGGGCGAGCGCTCTCTGAGAC 18736  
 QY 2118 ~~AlaAsn~~ValArgHisTyrPGLuLeuIleGlnGlnThrGluValGlyThrAlaTyrLeu 2138  
 Db 18737 AGCCAAAGGGGCACTGGAGCTGATTCACAGACAGAGGGGTGGCAAGCGCTGTCTCT 18796  
 QY 2138 uGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeu 2158  
 Db 18797 CCAGCACTATGAGGCTTACCGCAATGCGCTGGCCCAAGACATGGGGCACACTTACTAG 18856  
 QY 2158 rProPheThrIleValThrProAsnIle----- 2167  
 Db 18857 CCCCTTCACTTCGTCAAGCCCAACATTTGTATGAGCTGTGCTGGTGGGAGGGGTT 18916  
 QY 2167 ----- 2167  
 Db 18917 TGTGAGGAGAGTCCCGACAGAGCGGCTGTGCTGGGGTCTGCTGCTCAAGTCCG 18976  
 QY 2168 ----- 2168  
 Db 18977 ATCTGTGACATCCCTTCTTCTTACTCATCTCCGTAAGTCCGCTTGACAAAGGAACCTTG 19036  
 QY 2180 ~~IaGly~~AlaIleValLeuProArgTyrGlnAlaLeuArgGlyIleGlnProProAspLeuGluT 2200  
 Db 19037 CTGGGGCCMAAGCTCCCGCTTACAGAGGCGCTGTGGGGAGACAGCCCGGACCTTGAGA 19096  
 QY 2200 hThrValIleLeuProGluSerValPheArg----- 2210  
 Db 19097 CAACAGTATTTCTGCTGAGTCTGTCTTCAAG-AGTCAAGTGTGGCCATGTGATGAGTTG 19155  
 QY 2210 ----- 2210  
 Db 19156 GAGCTGAGACCCAGTGTCTGTGACAGTCCACAGAGAGGAGGCGCCAGCTAAGTGTGAC 19215  
 QY 2221 -----GluThrProProValValAlaProAlaGlyProGlyIleAlaGln 2226

```

Db      19216 AGTGTCCCTCCCAAGAACGCCCCCGTGTGACAGCCCGCAGAGCCCGGAGAGGCCAAG 19275
Qy      2226 1uProgluGluLeuAlaArgArgGluArgRhiSPProgluLeuSerGlnGluAlaV 2246
Db      19276 AGCCAGAGAGCTGGCAAGGAGACGACGAGCCCGAGCTGAGCAAGGGAGGCTG 19335
Qy      2246 aAlaSerVal11le1eYrArgThrLeuAlaGlyLeuLeuProH1SaenTYrASPPro 2266
Db      19336 TGGCCAGCGTCATCATCTACCGACCTGGCCGGGCTACTGGCTCACTAATGACCTG 19395
Qy      2266 splySaenSerLeu----- 2270
Db      19396 ACAAGCGCACTTGAAGGTACAGACTAGGGACAGGTGTGGTATGGGGTATGGGTCGGC 19455
Qy      2270 ----- 2270
Db      19456 GGTGAGTGTGAGGACATGAGAGGGGTGCGGGCGCTCTCCCAAGTCATGATGACCGGTG 19515
Qy      2271 -----ArgValProlySaenPro11e1eAnthrPro 2281
Db      19516 TGACTGTGACCTGACTGACCCCAATCAGAGTCCCCAAGCCCGATCATCAACACACC 19575
Qy      2282 ValValSer11eSerValH1SaenAPrgLgluLeuLeuProAlaAlaLeuAlaPro 2301
Db      19576 GTGGTGTGACATCAGCGCTCATGATGATGAGAGAGCTTCTGCCCCGGGCTGGAACAACC 19635
Qy      2302 ValThrValGlnPheArgLeuLeuGluThrGluGluArgThrArgPro11eCysValPhe 2321
Db      19636 GTCAAGGTGTGAGTTCGCCCTGTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19695
Qy      2322 TrpAsnH1Ser11eLeu----- 2327
Db      19696 TGGAACCATTCATCTGTGAGCTGTGACTGCTGCGCCCTCAGGCTTCGGGCTGAAA 19755
Qy      2327 ----- 2327
Db      19756 GTCCAGGCCCCGTCATGCTCACCCTCTGTCTCTCCCTGACCCCTGCTTCACACAG 19815
Qy      2328 ValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGlyValValPheArgAnGluSer 2347
Db      19816 GTCAAGTGTGACAGAGTGTGTGTGGCCAGAGGCTGTGAATCTCTTCGCAATGAGAGC 19875
Qy      2348 H1ValSerCysGlnCysAsnH1SmetThrSerPheAlaValLeuMetAPVal1SerArg 2367
Db      19876 CAGGTACGTGCGACGTGCAACACATGACGAGCTTGTGTCTCATGACGTTTCTCGG 19935
Qy      2368 ArgGlu----- 2369
Db      19936 CGGAGAGTGGGGCCCAAGGGGCACTGACAGAGCCGTGGGTGGGCAACCAAGGACGGGG 19995
Qy      2370 -----As 2370
Db      19996 CTGGGGTGTCAAGGTCTCGCCCTTCTTAATTCCTGGGCCCTGCGCACATCTGACAGA 20055
Qy      2370 nglyGlu11eLeuProLeuYrThrLeuThyYrValAlaLeuGlyValThrLeuAla 2390
Db      20056 TGGGGAGATCTCTGCACTGAAGAACAATGACATGCTGGCTCTGAGGTGCTCACTTGGCTGC 20115
Qy      2390 AlaLeuLeuThrPhePheLeuThrLeuLeuAlaGly11eLeuArgSerAsnGlnH1Sgl 2410
Db      20116 CTTTGTGCTACCTTCTTCTCTCACTCTGCGATCTGCGCTGCAACCAACACAG 20175
Qy      2410 Y11eArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGly11 2430
Db      20176 CATCCAGCGTAACCTGACAGCTGCTCGGGCTGGCTCACTGTCTCTCTCTGGGAAT 20235
Qy      2430 eAenGlnAlaAPLeuPro----- 2436
Db      20236 CAACCAAGGCTGACCTCCCTGTGAAGATGCTCTACTGCGCAGAACTGTCCCACTTCTC 20295
Qy      2436 ----- 2436

```

```

Db      20296 AGGCGGCTTCCCCAGGCCCCCACTGGCAACCCCTGTCTGTACATGAATCTAATAAG 20355
Qy      2436 ----- 2436
Db      20356 TGGCTATGTACAGACCTTGGCCAGGGTTTCTTCTTGTGTGCTCCCCGGGAATCCCCAG 20415
Qy      2436 ----- 2436
Db      20416 CACTGTGCTTGGGCCAAGGCTTCCGTGAAGAAGTTCCCAACCAAGGCCCTCTCAT 20475
Qy      2436 ----- 2436
Db      20476 GCCTGACCCGAAGCAGAGCTGTGCTGGGCGGGCCCCGTGTGCTGACCTGCCCCGGCC 20535
Qy      2437 -----PheAlaCysThrVal11eAla11eLeuLeuH1SPheLeuYrLeuCysT 2453
Db      20536 TGGGCCCTCATGTTGTGCTGACAGTCAATTCATCTGTGACATTCCTGATCCTGTGCA 20595
Qy      2453 hrPheSerTrpAlaLeuLeuGluAlaLeuH1SleuYrArgAlaLeuThrGluValArg 2473
Db      20596 CTTTTCTGGGCTGTGCTGAGAGGCTTGCACCTGTATCCGGGCACTCACTGAGTGGCGG 20655
Qy      2473 spValAsnThrGlyProMetArgPheTYrTYrMetLeuGly11TrpGlyValProAlaPhe 2493
Db      20656 ATGTCAACACCGGCCCATGTGCTTCTACTATGCTGTGGGCTGGGCGGTGCTGCTTCA 20715
Qy      2493 1eThr----- 2494
Db      20716 TCACAGATCTCCCAACCATTCCTGAGTCTTGGGGTCCACATCCCTGGGTCCACTTGT 20775
Qy      2494 ----- 2494
Db      20776 GCCATGTTCTTCCACCAATACAGGCCCTGAGGCCCAATCCCATGCCCCAGGCG 20835
Qy      2494 ----- 2494
Db      20836 CCTTATTCAGAGTGTCTCTGTATTAAACCACTGTGACCGGCAACCAAGCGTCA 20895
Qy      2494 ----- 2494
Db      20896 CTCCCTTATCTGGGGAGATCGTATGGGGCCGATGTGGGCAACCTTTTCCATGCTT 20955
Qy      2494 ----- 2494
Db      20956 CATGCTGGCCCTGTAGACCACTGCGCGAGCCCTACTTCCAGGCCCTCATCACCC 21015
Qy      2495 -----GlyLeuAlaValGlyLeuAPProglu 2504
Db      21016 CTCACCTGCTCCGCTGTCTCTCATGTCTCAGAGGCTA-GCGGTGGGCTTGAACCCGAGG 21074
Qy      2504 1YTrGlyAsnProAPPheCysTrpLeuSer11eYrAspThrLeu11eTrpSerPhe 2524
Db      21075 GCTAAGGGAACCTTGAATCTTGTGTGCTCTCATCTATGACACGCTCATCTGGAATTTT 21134
Qy      2524 1eGlyProValAlaPheAlaValSer----- 2532
Db      21135 CTGGCCCGGTGGCTTGTGCGCTCTCGGTGAGGTGACAGGTGGGTGGGTGTCACTGT 21194
Qy      2532 ----- 2532
Db      21195 GGCCTCTTGTGTCTCTCTGCTGTGAGGCGCGGTGTCTCAAGACTCCCTTTAGGA 21254
Qy      2532 ----- 2532
Db      21255 ACAGCTGAGGCAAGGGCCCTGTGGGTGGTGAAGCTGTTTCCCACTGAGCAC 21314
Qy      2533 -----MetSerValPheLeuTYr11eLeuAlaAlaAlaAla 2545
Db      21315 CCATGCTTGTGCCCCCTGCTGATGATGATGCTTCTCTTATCACTGTGCGGCGCGGCT 21374
Qy      2545 eCYeAlaAlaGlnArgGlnGlyPheGly11eYrGly11e----- 2557
Db      21375 CCTGTGCTGCCAGGCGGAGGCTTTGAGAAAGAGTCTGTGTGATATAGGCTTGG 21434

```

QY 2557 ----- 2557  
 Db 21435 GTGCTGGGCGATGGGACAGACAGATGGAGGTCTCATGGCCGGACTCATGGCTGT 21494  
 QY 2558 -----ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT 2575  
 Db 21495 CCTATCCCGAC-TCGGGCGTGCAGCCCTCTCGCCCTCGCTGCTGCAGCGGCA 21553  
 QY 2575 hTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheA 2595  
 Db 21554 CCGTGCCTGGGCACTGCCTCTCTGTCAACAGACACCTCTCTTCCACTCACTCTTTG 21613  
 QY 2595 IaThrCysAsnCysIle----- 2600  
 Db 21614 CTACCTGCATTTGCAATCCAGTACTGCGCCAGCCTGTGAGAGAGAGACCTGGGCT 21673  
 QY 2601 -----GlnGlyProPheIleP 2606  
 Db 21674 GTGATGCTGTAATATGACAGACCGTGTGCTGCTTGGCTGTGCAGGGCCCTTCATCT 21733  
 QY 2606 heLeuSerTyrValValLeuSerTyrGluValArgLysAlaLeuLysLeuAlaCysSerA 2626  
 Db 21734 TCCTCTCTATGTGGTGTGCTTACAGAGAGAGTCCGAAAGCACTCAAGCTTGCCTGAGCC 21793  
 QY 2626 rGlyAsProSerProAspProAlaLeuThrThrLysSerThrLeuThrSer----- 2642  
 Db 21794 GGAAGCCCGAGCCCTGACCTGTCTGTGACCAAGTCCACCTGACCTTGGTGAAGAGAC 21853  
 QY 2642 ----- 2642  
 Db 21854 CAGGGGTCTCAGAGAGCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21913  
 QY 2642 ----- 2642  
 Db 21914 TTGGGCCCACTCCCTTCTCTTCTTCATCCCTCTCTGAAAGTGAAGGAGGAGTG 21973  
 QY 2642 ----- 2642  
 Db 21974 AAATGCTGTGTGGGCTGGGAGAGAGATGTTGAAAGGGTGTAGATTGACTTGA 22033  
 QY 2642 ----- 2642  
 Db 22034 GGGAGATGGGCTGAGACAGATATCAAGTAATAAGTCTTGACCTGAGATCTCTCA 22093  
 QY 2642 ----- 2642  
 Db 22094 TGAATTTAGTAATTTTCTGTCTCTCCAAAGTGTCTCAGAGAGACATTTCTGTCTCAT 22153  
 QY 2642 ----- 2642  
 Db 22154 CTTTTCAGAGCCATTATATAGTAGTCCAGGATCCCTGGGGGAGAGAGGGGTAGACA 22213  
 QY 2642 ----- 2642  
 Db 22214 GCTGTGGCAGAGCTGTGACCACTGCAGGCTCAATATGGGAGCAGCGAGAGAGAGGCG 22273  
 QY 2642 ----- 2642  
 Db 22274 CCTCTCTTTCAGACCTGAGCATGTGGAAGTGAACCTTAGGGCAAGTTCCTCCACCTTC 22333  
 QY 2642 ----- 2642  
 Db 22334 CTTCTTCCAGCCTGGGGCAGACCATCCCACTCCCACTTACTGACTCTGTCTCCCTG 22393  
 QY 2643 -----SerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyA 2661  
 Db 22394 CCTATGCTTCAACACTGCCCCAGCCCTTAGCAGATGGGGGCTGTACCAAGCCTTAGCGAG 22453  
 QY 2661 sPserAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIleP 2681  
 Db 22454 ACTGGCGGGGCTCTGTGACAGACCAAGTGTGCTGGGGAAGGTACGCCAGCTACATCC 22513

QY 2681 rPheLeuLeu----- 2684  
 Db 22514 CTTTCTTCTGAGGTGAATCCCGAGATGGAGGTGGAGAGAGAGAGAGAGAGAGAGAG 22573  
 QY 2684 ----- 2684  
 Db 22574 CATGTGAGCCAGCCAGCCAGCTGTGGAGTTGAGAGACACACTGTGGCTGAGT 22633  
 QY 2684 ----- 2684  
 Db 22634 GGGGCGAGCTTGATTGAAGCTGTAAAGGACCAACAGCAGAGAACAGATCCAGGGG 22693  
 QY 2684 ----- 2684  
 Db 22694 AGAGAGAGACTGGGACCTTGGGAGAGGGGCGAGGCTGACCCCTTCAGATGTCTATC 22753  
 QY 2685 -----ArgGlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLysProG 2703  
 Db 22754 TTCCTAGGAGAGAGTCCGCACTGAACTCTGGCAAGGGCCCTTGGCTTGGGGGATCCAG 22813  
 QY 2703 LysLeuPheLeuGlnGlyGlnAspGlnGlnIleAsp----- 2715  
 Db 22814 GCAAGCTTCTCTGAGAGGTAAAGACAGACAGATGC-TGAGACAGAACCTTGGCCA 22872  
 QY 2715 ----- 2715  
 Db 22873 CCCAGAGAGGAGTTGGCTGCTTTTACTGAAGGTGGGTGAGAGGTGCTGGGCTGCTGT 22932  
 QY 2716 -----ProAspThrAspSerAspSerAspLeuSerLeuGlnA 2728  
 Db 22933 GATCTCTCCCTGGCTCTAGATCTTGAACGAGACTCCGACAGTGAACCTGTCTTAGAAG 22992  
 QY 2728 sPAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlyGlnGlnGlnGln 2748  
 Db 22993 ACGACAGAGTGGCTCTTAAGCTCTTACCACTATCAAGACGTAGAGAGAGAGAGAGAG 23052  
 QY 2748 LngLngLngLngLnaAlaPheProGlyGlnGlnGlyTyrPAspSerLeuLeuGlyProG 2768  
 Db 23053 AG 23112  
 QY 2768 LysAlaGlnArgLeuProLeuHisSerThrProLys----- 2779  
 Db 23113 GAGCAGAGAGACTGCCCTGTGACAGTACTCCAAAGGTGGGCGAGCACTGGGGCTGTGGC 23172  
 QY 2779 ----- 2779  
 Db 23173 CTTTGGGGCGAGTGGAGAGACAGTGGGCTTGTGGGAAGCAGAGCTGGGCTGTGT 23232  
 QY 2780 -----AspGlyG 2782  
 Db 23233 GGCTCTGCTGATCTGCCCTGTGGGCTCATCTTCTTCCCAACAGATGGGG 23292  
 QY 2782 LysProGlyProGlyLysAlaProThrProGlyLysAspPheGlyThrThrAlaLysGlnLys 2802  
 Db 23293 GCCAGGGGCTGGAG 23352  
 QY 2802 erGlyAsnGlyAlaProGlyGlnGlnArgLeuArgGlnAspAlaLeuSerArgGlnG 2822  
 Db 23353 GTGGCAAGGGGCGCTCGAGAGAGGCGCTGGGAGAGATGAGATGCCCTGTCTGAGAGG 23412  
 QY 2822 LysLeuLeuGlyProLeuProGlyLysSerAlaGlnProHisLys----- 2836  
 Db 23413 GGTCCCTAGGCGCCCTTCAAGGCTCTTGTGCCAGCTTCAAAAGTGAAGTGGGACCC 23472  
 QY 2836 ----- 2836  
 Db 23473 CCAAGTCCAGAGCTCCCTATGTACAGAGCCTCATACTCATCTTCTGTGGCCGAGCC 23532  
 QY 2837 -----GlyTyrLeuLysLysCysLeuProThrIleSerG 2849  
 Db 23533 TCACAGCCCGGCGCCGAGCCAGAGGATCTTAAAGAGAGAGTGTCTGCCACCATCAGCG 23592  
 QY 2849 LysSerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSerArgGlySers 2869

Db 23593 AGAAGACAGCCTCTCGCGCTCCCTCGAGCATACAGAGGTCTTCCCGGGCTCCT 23652  
 Qy 2869 enlaserlgllyserlgllylproprobaargproprobaarglinsertleug 2889  
 Db 23653 CCGCTAGTGGGGAGCGGGAGAGCCCTCCCTCCCGCCAGCGCCCGGCAAGCTCC 23712  
 Qy 2889 lnglgluInleAaenlglYalmeCProIlealmeSerleleYalaglYThrValaspC 2909  
 Db 23713 AGAGGACGCTGAACGGGGCTATGCCATCCGATGAGCATCAAGGAGGACCGGTGATG 23772  
 Qy 2909 luhapserlgllyserglu 2915  
 Db 23773 AGGACTGCTGAGGCTCCGAG 23792

RESULT 9  
 AB065955  
 LOCUS 24370 bp DNA linear PRI 23-JUL-2002  
 DEFINITION Homo sapiens gene for seven transmembrane helix receptor, complete cds, isolate:CBRC7TM\_518.  
 ACCESSION AB065955  
 VERSION AB065955.1 GI:21929187  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 Suwa, M., Sato, T., Okouchi, I., Arita, M., Putani, K., Matsumoto, S., Teutemul, S., Aburatani, H., Abai, K. and Akiyama, Y.  
 Genome-wide discovery and analysis of human seven transmembrane helix receptor genes  
 Unpublished  
 2 (bases 1 to 24370)  
 Suwa, M.  
 Direct Submission  
 Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan (E-mail:m-suwa@aist.go.jp, URL:http://www.cbrc.jp/, Tel:81-3-3599-8080, Fax:81-3-3599-8081)

COMMENT This sequence is a seven transmembrane helix receptor candidate predicted from the whole human genome sequences using our automated system that contains programs of gene finding (GeneDecoder), sequence search, motif-domain assignment and transmembrane helix prediction.  
 And the sequence is submitted by the collaborative project between [Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)] and [Genome Science Division, Research Center for Advanced Science and Technology (RCAST), University of Tokyo].  
 Location/Qualifiers  
 1. 24370  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /isolate="CBRC7TM\_518"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 join(201..3510,8553..9200,11163..11385,11634..11749,11929..12017,12408..12565,12927..13090,13263..13388,13732..13898,14199..14486,14575..14749,14989..15130,15312..15432,15856..16039,16226..16327,17669..17763,17972..18181,18702..18885,19001..19129,19231..19411,19545..19712,19815..19941,20054..20254,20547..20721,21048..21160,21338..21417,21507..21633,21721..21843,22399..22526,22761..22849,22954..23148,23287..23457,23557..23791,24143..24170)  
 /codon\_start=1  
 /evidence=not experimental  
 /product="seven transmembrane helix receptor"  
 /protein\_id="BAC06168.1"  
 /db\_xref="GI:21929188"  
 /translation="MRSPATGVPLPTPPPLLLLLLLPPLPLGGVGVGRSLGSRG

RSSGACAPMGWLCPSSASNLMLYTRCRDAGTELTGHLVPHDGLRVWCESSEAHIP  
 LPPAPGPMSCRLIGIGHLSPOKLTLPREHPCGLKAPRLCSGCAIDDAKAPGABG  
 RSPRESLGRRRKRNNTAPOROPSVQVATVENDPAGPVRASLRALDDKDEKALRY  
 TMDALPDRSNQPSLDPVTGAVTTAAELIDETKSTHPRRTAODHMPRSALATLT  
 ILVDTNDHDPVFEQOETKESLRENLREGLVTRADGDAPRANLTLRLBSGG  
 SPSEVPEIDPSRGVTRGPDVREVEVQLVTEASDGDGDPGAPLTAALVLEVD  
 NDMAPOFSEKRYVQVRDREDVTPGAVLQSGDRDKGNAVWHSIMGNARQGYLD  
 AOTGLDVAVSLDPEVETKETTLRVAAQGGGPPPLSNVGLVTVQVLDINDAPLIVST  
 PROATVLESVPLGVLVHVOALDADGASNRVLEVLACVGHDPPTTINNGVMISVA  
 ELDRERUDPSRGVAVRDCGPALTAASVAVTVLDVNDNPTPROPTVAVLNDA  
 VETSIVVTSAYDRDASHVITTOITSGTRNRPSTTSQSGGCVSLALPLDYKLEQV  
 LAVTASDGTRODPAQIVNVTDANTHRVPSSTHATVAVNEDRPAGTVLISADSD  
 TGENARITYFMEDSIPQRIIDAGTAVTQALREVDVSTYLAITADNDLPQSDT  
 TYLILVNDVNDNAPOFLDRSIOGSVDYDPEFTSVLQISATDRSGNLGRVFTVQ  
 GDDGDGDFIVSSTSGIVRTLRIDRENEAOVYLRAYADKMPPARVTEVAVTVD  
 NDNPVPRDREDVDVRENSPGLAVATVATDDEGNAOMQIVGNIPEVOLD  
 IPSGELTALVDIDREPREVILVTOATSAPLVSRATVAVRLDRDNPVAGNEFTLP  
 NNVTNRSSSPFGAIGRPVAPHDPIISLTYSPERGBLSVLLNASTGELKLSRAL  
 DNMRLEAIMSVLSDVGVSVTAQCALVITITBMLTHSITLRLDMSPERFLSPL  
 GLFIOVAVATLATPDHVVNVVNOEDTAPGGHILNVLSVQPPGPGGPFLESD  
 LOERLYNRSLLTASAOVLPEDNICLARECEYMRCSVLRDSSAPFLASSVL  
 FRPIHPVGLRCRCPGFGPYCETEVNLCYSPGPHGRCSRRGGVTCICRDGYTG  
 EHCIVASASGCTPVCNKGCTCVALLVGRKDCDPSDFEPYQVYTRSPRPSPI  
 TFRGLRORPHFTLSTFAKEDGLLNGRPNRHDVVALEIOEQVLTFSAGEST  
 TTVSPFVGVSDGQMTVOLKYNKPLLGQGLPQGSBQVAVVTVQDCDTVALR  
 FGSVGVNTSCAQTQGSKSLDITGLIGVLDLPSBSPVNRQFVGCMKRLQVD  
 SSIIMADPIANGTVPGCPAKNCVNDTCNGTCVNMOPAFCECPLAGGKSCA  
 OMANPOHRLSSIVAMHGLSPISOPVYLSIMPTRRADGVLTALIRGSEITLQ  
 REGHMLSVBEGGLQASSRLRERGANGDHMHQALAGAAGSPHAILSPYQORA  
 BENLPRHLGHLSITVIGLIRPAGVARGRGGLQGRVSDTEGNSLDPBSGES  
 INVEQCSLPDPCDSNCPANVSCSNDSDSCSDPYGNCNVCNDLNEQHSY  
 CTRKSPABGYTCBPNVLAGPYCETRIDPCPRMGHPTGCPNCVDSKGFDCDN  
 KTSGECHKENHYPSPGSPCLCDYPTGSRVCDPDDGCPKPGVIRGQCDRG  
 NPFAVTNGCEVNVDSCPRAIEAGIMPRPRFGAPAAFCPSFGVAVRCDHRG  
 WLPPMFLNCTSTTSEELKPRARLQRNSGLDGSQCLALIRRAOHTGYRCSGV  
 KYAVOLATRLAHBSTQKFGSLATQDHFLENLVSALILDTLANKAMELIOQTEG  
 GTAMLLQHYEAVASLAQNMRTYLSPTITVPIVVISVARDKNGAGAKLIRREAL  
 ROEOPDLETTLVLPESVRETPVRPAGPEAOPEBLARQRBHELSGEAVAS  
 VIITTLGLAPLPHNDPDKRSILRVKRIPLINPVVISVHDDEQLPALDSEVQF  
 RLLEERERKPLCVFMNHSILVSGTGMASRCVFRFNSHSSCCNMHTFAVMD  
 VSRNRNGRIPLKTLTYVALAVTVALAILTPPLTLITRSHONGIRBNITAAALGA  
 QVPLTGINQADLPACTVITAILHLPLTICTPSMALBRLHLYRLALTEBDVNTGPMK  
 FTYMLGWPVAPITLAVGLDEGSGNDPCLVITVTLNFPAGVAVASVSLFYAT  
 IIAAASCAQORQGFKEKQPVSGQPSVALLSLATYMLLANSVDLFAVMSVPLT  
 CNICIGPPIFLSVYLSEKVRKALKALCKRSKSPALATKSTLTSSVPCPYADGR  
 LYQPGDSAGLSHSTRSGKSPSYIPLFLRESLANPCQPGGGLDPSLPLEGDD  
 OHDPTDSNDLSLDDOSGYASVHSSDSSEEREBEREAAPFGQGMDSLGPAB  
 RLPLHSTPDGSPRGKAPMBGDRTYAKESGNGNAPERLALENDAPLSRSLSGLPL  
 GSAOPHKLTLKKCLPTTISKSLRLRLEOCTSSRSSASBSRGSGPPRPAPRQ  
 SLQEQINGMPLAMSIKQGVDESSGFLEFNLH"

BASE COUNT 4543 a 7241 c 7458 g 5128 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 24370  
 Score: 12339.00 Matches: 2907  
 Percent Similarity: 36.92% Conservative: 8  
 Best Local Similarity: 36.92% Mismatches: 0  
 Query Match: 79.38% Indels: 4967  
 DB: 9 Gaps: 32

US-09-916-849A-3 (1-2923) x AB065955 (1-24370)

Qy 1 MetArgSerProAlaThrGlyValProLeuProTrpProProProLeuLeuLeu 20  
 Db 201 ATCCGAGCCCGGACCGGCTCCCTCCCAAGCGCCCGCCGCTGCTGCTCG 260  
 Qy 21 LeuLeuLeuLeuProProProLeuLeuGlyaaPGLNalGlyProCysaArgSerleu 40  
 Db 261 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320  
 Qy 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCysProSer 60

Db 321 GGSTCCAGGGGACGAGGCTCTTGCGGGGCGCTGGCCCCCATGGGCTGGCTCTGTCCATCC 380  
Qy 61 SerAlaSerAsnLeuThrLeuThrLeuThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
Db 381 TCAGGCTCCAGACTCTGGGCTCTACACAGCCGCTGCAGGGATGCGGGCATGTGAGTCACT 440  
Qy 81 GlyHisLeuValProHisAspGlyLeuArgValThrCysProGluSerGluAlaHis 100  
Db 441 GGCCACTGGATCCCAACACAGATGAGGCTGGAGGGTTGGTGTCAAAATCCAGAGGCCAT 500  
Qy 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
Db 501 ATTCCTCCACACAGACTCTGAGAGGCTGCCCTGAGGCTGTGCTCTCTGGGCATTGGA 560  
Qy 121 GlyHisLeuSerProGluGlyValLeuThrLeuProGluGluHisProCysLeuValAla 140  
Db 561 GGCCACCTTCTCCCAACAGGGCAAGCTCACACTGCCAGAGAGACCCGATTAAGGCT 620  
Qy 141 ProArgLeuArgCysGlnSerCysIleuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
Db 621 CCAAGGCTCAGATGACAGATCTGCAAGCTGCACAGGCCCGGGGCTCAGGGCAGGGGAA 680  
Qy 161 ArgSerProGluGluSerLeuGlyValArgArgValAspAsnValAsnThrAlaProGln 180  
Db 681 AGGTCAACGAGAAGTCCCTGGGTGGCGTCCGAAAGAAATGTAAATCAAGCCCCCAG 740  
Qy 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
Db 741 TTCACGCCCCCAAGCTACAGAGCCACAGTCCGAGAAACAGCAGCAGCAGCAGCCCTGTT 800  
Qy 201 AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyArgLeuGluTyrThrMet 220  
Db 801 GCATCCCTGAGGGCCATTCACCCGAGCAGAGGTGAGGAGGTGAGTCACTGAGTGAACCAAG 860  
Qy 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
Db 861 GATGCCCTCTTGATAGCCGCTCCAAACAGTCTTCTCTCCCTGAGCCAGTCACTGGTGA 920  
Qy 241 ValThrThrAlaGluGluLeuAspArgGluThrIleValPheArgValThr 260  
Db 921 GTAAACCAACGAGGAGCTGATCGTGAACCAAGAGCACCAAGTCTTCAAGGGTCAAG 980  
Qy 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
Db 981 GGCGAGAGACAGGAGCAGTCCCGACAGAGTGCCTGGCTACACTCACTTGTGTACT 1040  
Qy 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrIleGluSerLeuArgGlu 300  
Db 1041 GACACCAACGACATGACCTCTGTTCAGACAGCAGAGTACAGAGAGCCTCAGGGAG 1100  
Qy 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
Db 1101 AACCTGGAGGTGGCTATAGGTGCTCACTGTCAAGGCAACGAGTGTATGCCCTCCC 1160  
Qy 321 AsnAlaAsnIleLeuThrArgLeuLeuGlySerGlyIleSerProSerGluValPhe 340  
Db 1161 AATGCCAATATCTGATCCGCTGCTGAGGGGTCTGGGGGCAAGCCCTCTAAGTCTTT 1220  
Qy 341 GluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
Db 1221 GAGATGAGCCCTCGCTCGGGGTGATCCGAACCGTGGCTCTGTGATGGGAAAGGTG 1280  
Qy 361 GluSerTyrGluLeuThrValGluAlaSerAspGlnIleArgAspProGlyProArgSer 380  
Db 1281 GATCTCTACACAGCTGCGGTAGAGGCAAGTGAACAGGGTCGGGACCCGGGTCTCTGAGT 1340  
Qy 381 ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPheSer 400  
Db 1341 ACCACAGCCGCTGTTCTTCTGTGAGAGTGAACATGATTAATGCCCCCACTTAAAT 1400  
Qy 401 GluIleArgTyrValValGluValArgGluAspValThrProGlyAlaProValLeuArg 420  
Db 1401 GAGAAAGCGCTATGTGGTCCAGGTGAGGAGGATGTGACTCCAGGGGCCCACTATCTCCA 1460

Qy 421 ValThrAlaSerAspArgAspIleGlySerAsnAlaValAlaHisTyrSerIleMetSer 440  
Db 1461 GTCAAGCCTGGATCGACAGAGGGAGCAATGCCGTGGTGCATATATGATCATTAAGT 1520  
Qy 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValLeuAspValIleSer 460  
Db 1521 GGCAATGCTCGGGAGCAAGTATTATCTGATGCCAGACTGGAGCTGTGATGTGTGAGC 1580  
Qy 461 ProLeuAspTyrGluThrThrArgGluTyrThrLeuArgValArgAlaGlnAspGly 480  
Db 1581 CCTCTGACTATGAGACACACAGAGTACACCTTACGGGTGCGAGCACAGATGGTGC 1640  
Qy 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1641 CGTCCCCACTCTTAATGTCTCTGGCTTGGTGAAGTACAGGTCTGTGATCAACAGC 1700  
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
Db 1701 AATGCCCATCTTCTGTGAGCACCCCTTCCAGGCTACTGTCTGTGAGAGTGTCCCTTA 1760  
Qy 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
Db 1761 GGCTACCTGGTCTTCATGTCCAGGCTATCCAGCTGATGCTGGTGAATGCCCTG 1820  
Qy 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
Db 1821 GAATACCGCTTGTGGTGGGTGGACATGATCTTCCCTTCAACATCAACATGGCACAGGC 1880  
Qy 561 TrpIleSerValAlaAlaGluLeuAspArgGluValAspPheTyrSerPheGlyVal 580  
Db 1881 TGGATCTGTGGCTGTGAACTGACCGGAGAGAGTGTATTTCTACAGCTTGGGGTA 1940  
Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThrVal 600  
Db 1941 GAGCTCAGACCATGGACCTCCAGCACTCACTGCTGGCCAGTGTACAGCTGACTTC 2000  
Qy 601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
Db 2001 CTGGATGTCACACGACAAATCCAACTTATCCCAACAGAGTACACAGTGGGCTCAT 2060  
Qy 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
Db 2061 GAGATGAGCTGTGGGACCAAGGTGTGACGGTGTGACGTGTGACCGGTATGCTCAT 2120  
Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
Db 2121 AGTGTCACTCACTTACAGATCAACAGTGGCAATCTGAAACCGCTTCCATCAACAGC 2180  
Qy 661 GlnSerGlyGlyIleLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluArgGln 680  
Db 2181 CAAAGTGGTGGGTGTGATCCCTTGCCTGCACTGAGCTCAAACTTGAAGGGCAG 2240  
Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleValVal 700  
Db 2241 TATGTGTGGCTGTACCGCTTCCAGTGGACCTGGGAGGACAGAGCAATGTGTGTG 2300  
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
Db 2301 AATGTCAACGACGCAACACCACTGCTCTTCTTCAAGCTCCCACTAATCAAGTAAAT 2360  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
Db 2361 GTTATATGAGACCGGCGCGGAGGACCAAGGTGTGTGATCAAGCCCAAGATGAGGAC 2420  
Qy 741 ThrGlyGluAsnAlaArgIleThrThrPheSerGluAspSerIleProGlnPheArgIle 760  
Db 2421 AAGGTGAGATGCGCGATCACTACTCATGAGAGCAAGCAATCCCAATTCGGCATTC 2480  
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
Db 2481 GATGAGACAGGGGGGTGTCAACCAAGCTAGCTGAGCTATGAAGCAATAGTGTCT 2540



QY 781 TTTTThleuAla1leThra1aAArgAspAsnGly1leProGlnYSerAspThrThrTyT 800  
DB 2541 TAAACCTGGCCATTACTGCTCGGAGCAATGGCAATCCCGAAGATCCGACCACTTAC 2600  
QY 801 LeuGln1leleuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyT 820  
DB 2601 CTGGAGATCTCGGTGAAACGACGTGAATGACAAATGCCCTCAGATTCTCGCAGACTCTTAC 2660  
QY 821 GlnGlySerValTyTyrGluAspValProProPheThrSerValLeuGln1leSerAlaThr 840  
DB 2661 CAGGGCAGTGTCTATGAGATGAGCCACCTTCACTAGCGTCTCTGACAGATCTCAGCACT 2720  
QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyTThrPheGlnGly1leAspAspGly 860  
DB 2721 GATCGATATTCTGACTTAATGACAGGCTTCTTACACTTCCAGAGAGCGACGATGGA 2780  
QY 861 AspGlyAspPhe1leValGlySerThrSerGly1leValArgThrLeuArgGlyLeuAsp 880  
DB 2781 GACGGTGACTTATTTGTGATGTCACGTACAGCATCGTGCACGCTACCGAGGCTGGAT 2840  
QY 881 ArgGlyAsnValAlaGlnTyTValLeuArgAlaTyTAlaValAspLysGlyMetProPro 900  
DB 2841 CAGAGAACCTGGCCCGCATGATGTTGTCGGGCAATGCAAGTGAACAGAGGAGATGCCCA 2900  
QY 901 AlaArgThrProMetGlyValThrValThrValLeuAspValAsnAspAsnProProVal 920  
DB 2901 GCCCGCACACTTATGGAAGTGAAGTCACTGTGTGGAATGTAATGACATCCCTGTC 2960  
QY 921 PheGlnGluAspGluPheAspValPheValGlyGluAsnSerPro1leGlyLeuAlaVal 940  
DB 2961 TTTGACACAGATGATGATTTGATGTTGTGGAAGAGAACGCCCATTTGGGCTAGCCGTG 3020  
QY 941 AlaArgValThrAlaThrAspProAspGlyGly1leThrAsnAlaGln1leMetTyTyrGln1le 960  
DB 3021 GCCCGGCTCACACCACTGACCCCGATGAAGGACCAATGCCAATTAATATGCCAGATT 3080  
QY 961 ValGlyGlyAsn1leProGlyValPheGlnLeuAsp1lePheSerGly1leLeuThrAla 980  
DB 3081 GTGAGAGGCAACCTCGTAGGTCTTTCAGCTGGAACATCTTCCGGGGAGCTGACAGCC 3140  
QY 981 LeuValAspLeuAspTyTyrGluAspArgProGluTyTValLeuVal1leGlnAlaThrSer 1000  
DB 3141 CTGTGACTTAACATACGAGACCGGCTGAGTACGTCTGTGCATCCAGGCAAGTCA 3200  
QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
DB 3201 GCTCTCTGTGAGCGGCTACAGTCCAGCTCCGCTCTTGAACGGCAATGACACCA 3260  
QY 1021 ProValLeuGlyAsnPheGln1leLeuPheAsnAsnTyTValThrAsnArgSerSer 1040  
DB 3261 CCAAGTCTGGGCAACTTGGATCTTTTCAACACTATGTCAACAATGCTCAACAGC 3320  
QY 1041 PheProGly1leGlyAla1leGlyArgValProAlaHisAspProAsp1leSerAspSerLeu 1060  
DB 3321 TTCCTGTGGGTGCATTTGGCCAGTACCTGCCCAGTACCCTAATATCTCAGATAGTCTG 3380  
QY 1061 ThrTyTSerPheGlnGluArgGlyAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
DB 3381 ACTTACAGCTTTAGAGGGGAAATGAATCAAGCTGTCTGTCTCAATGCTCAAGGGT 3440  
QY 1081 GlnLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAla1leMetSerVal 1100  
DB 3441 GACCTGAAGCTTAAGCCGCGCATGACACACACCGGCTCTGGAGGCGCATGAGACGTG 3500  
QY 1101 LeuValSer----- 1103  
DB 3501 CTGGTGTTCAGGTAAAGAGGCCCAAGTGGCGCTGGGGTGGGGTAACTCGCGGGAGT 3559  
QY 1103 ----- 1103  
DB 3560 GTCTGGGCAAGCACTGAGAGGTAGGGTGCATCCAGGAAGCAAGTACAGATCCACTCCCT 3619  
QY 1103 ----- 1103

DB 3620 GCCCAGTGTCTGGCACAGAGCAGAGGGGCAAGACCAGCTTTGGGAAGAGCCCAAGAA 3679  
QY 1103 ----- 1103  
DB 3680 CCTGGCGGCTGGTGCAGGCCCCCACTTCTGTGGCTGGGCTGGGCTGCTGTGGTTGC 3739  
QY 1103 ----- 1103  
DB 3740 TCCCCCACTGTCTACAGACCTGTGACGTGATCGGGAAAGTGTGTAGAGCTTCCCCGCG 3799  
QY 1103 ----- 1103  
DB 3800 TATGGCTGCAGAAATTGTGAAAAAAGGCGGTGAGAGCTTGGGGGCGAGAGAGCTGG 3859  
QY 1103 ----- 1103  
DB 3860 GGAAGGCGAGAGAGACTGTGGCGCAGAGAGAGAGGCGAGGCCACTTCACCCC 3919  
QY 1103 ----- 1103  
DB 3920 CATCAAGATATTCTGACGTCTCACACAGCTTTGGCAGAGCTGGGGACACAGGAGAC 3979  
QY 1103 ----- 1103  
DB 3980 AGGCTGCGCCACCCAGGGGTCAAGAACCTGTAGTCTTCTTGGCCCTTTTGTCTCC 4039  
QY 1103 ----- 1103  
DB 4040 TGGTAGCTGCAAGATGCGGTGACAGGAACTTTGTAGCGCATTTTTCCTTCACAG 4099  
QY 1103 ----- 1103  
DB 4100 CAGAAAGGCTGGGCTGTGCTCCCTTCAAGACTCCACAGAAAGTGAAGGCCAGAGTTC 4159  
QY 1103 ----- 1103  
DB 4160 ATTGTCTGCCAAGACAGACCCCACTTTCAGAGTCCGTGTTCCGGTCTGGCTTGGAGAG 4219  
QY 1103 ----- 1103  
DB 4220 GGGCTGCAATGAAGCTGGGGGGCAGGCTGAGGCTGTGAGGCCCCCTTGCACCTCATG 4279  
QY 1103 ----- 1103  
DB 4280 GGTGTCTTAGGTTGACAGGCTTCTTCTGTAGCTGTGACAGGCCAAAGCTCCAG 4339  
QY 1103 ----- 1103  
DB 4340 GAATGAAGTGAAGGACAGACAGACGCAATGACTGTGACTTTGAAGGACAGAGACC 4399  
QY 1103 ----- 1103  
DB 4400 CAAAGCACTTATCTGAGAGAGAAACAGAGTCCAGAGGCGTGGAGGCTGGCCGAACC 4459  
QY 1103 ----- 1103  
DB 4460 AGGCTGGGCAATTCCTGTGCTTGGCAAGCACTCAGCTTACAGACTCCCTGAGACAGAG 4519  
QY 1103 ----- 1103  
DB 4520 GGTGGGGAGAGAGATTGAGAGGGGTTGTGACAGTATCTGTCTTTTATACAGCTG 4579  
QY 1103 ----- 1103  
DB 4580 TGGGAGCCAGAGGTAGCAGCAGGGAAGAACCAACCAATTGTGTAGTGTCTTCCCC 4639  
QY 1103 ----- 1103  
DB 4640 AAGCTGGGAACACTGAGACCTCATGAGGGAGTGTGTAGTCAACCCAGGGGTGG 4699  
QY 1103 ----- 1103

Db 4700 ATCAGGTCAATCCCTTAGGCAAGCTTCAATCATCAGGGAAGGAGAAAGAGA 4759  
 QY 1103 ----- 1103  
 Db 4760 GCGTGGCAGGCCCCCTACTCCACCCAGGAGATGTGCTGAAGGGAGAGGGGTANG 4819  
 QY 1103 ----- 1103  
 Db 4820 CCAGAGTCAGGGCGTGGCTCACTCAGGGTCTTGTGTGACAGATGCTCCCTGCCCCAC 4879  
 QY 1103 ----- 1103  
 Db 4880 CTTCAAGTAGCCACAGAACCAAGAGAGAGCTTTTCAAGGACTGTGGCTGAGTT 4939  
 QY 1103 ----- 1103  
 Db 4940 CTGAGCTCTGAGACTTCAAGTTTCCCTCAGAGAGCGGAAGAGGAGGCTGAGCTCC 4999  
 QY 1103 ----- 1103  
 Db 5000 TGTGTCTCACTGTGCTGGGCACTTGACAGAGTGGAGAGAGTGGAGAGAGGCGC 5059  
 QY 1103 ----- 1103  
 Db 5060 TCTCCCTCTGTCTCCACACCCAGCTGCTTTGTGTAGCGTGGCAGGCACTGGCAGA 5119  
 QY 1103 ----- 1103  
 Db 5120 GAAGGCCCCCATTAAGCTTGGTCTCCACAGGGGAGAGCCCTAGACTTGGTTCC 5179  
 QY 1103 ----- 1103  
 Db 5180 TAGTCTAGAGAGGCTCGGGGACATCCAGAGACGGCTCTCTGGCCGCGAGCTTGG 5239  
 QY 1103 ----- 1103  
 Db 5240 CCCCACCTCTCCCTTCGGGCGCTGACCTTGGAGGAGTACACAGGCAACCCCAAGTT 5299  
 QY 1103 ----- 1103  
 Db 5300 CAGGAGCGCGGGTGAAGGGAGAGTACTCCACTGCCCCCTTTGTTTCTCTCTGTGT 5359  
 QY 1103 ----- 1103  
 Db 5360 TCTTTGTCTCTTTCTTTCCCGCCCTGACCATGCGAGCCCTGACACTGCTCTTC 5419  
 QY 1103 ----- 1103  
 Db 5420 TCCGGTGCCTCTCCCTGTCTCTGGGGCTCTCCGCTCTCCACTTGGCCATTGCTGGCT 5479  
 QY 1103 ----- 1103  
 Db 5480 CCTCTCTGAGGCTTGTGCTTCCCTGTGTTGTGTCCTGTTGTGCTTCCACTCTCTTC 5539  
 QY 1103 ----- 1103  
 Db 5540 CCTATGCTGAGCTGCCAGCTGGGCCAATGCTCCCTGAGGCTGAATCTCTTC 5599  
 QY 1103 ----- 1103  
 Db 5600 GCGTCTTCTCTCTGTCTGACAGACAGAGACTCTGTCTCATGAGTCTGGCAGCTCCC 5659  
 QY 1103 ----- 1103  
 Db 5660 AGAAGAGAGAGCCCTCTTGGCTCCAGTCTCCGCTGGACATCTGCCAGGGAGAGAG 5719  
 QY 1103 ----- 1103  
 Db 5720 GCTGGAGGCTGACCGGAGCTTGACATTTTCCAGGTGTGCTGCTAAGCTTGGACAC 5779  
 QY 1103 ----- 1103  
 Db 5780 AGGTGGCATTTCTCTGGGGTCAAGGCGCATCCAGCTGTTCTTTCTGTCCAGAGAGCA 5839

QY 1103 ----- 1103  
 Db 5840 GCTCAGGAAGTACTCCCAACCCCTGTTCCCATCTGCTTTGCACTGAGGGGTGACC 5899  
 QY 1103 ----- 1103  
 Db 5900 TAGTCAAGCAGGGAGGTTGATATAGAAAGAGAGACAGTGGGTTGACTTGGGGA 5959  
 QY 1103 ----- 1103  
 Db 5960 TGGGGACACACGACAGGGGGCAGGTGGCCAAGATGTGTGACCATGTGGAGCAACA 6019  
 QY 1103 ----- 1103  
 Db 6020 GAGACATGTGGAGAGGGGCCAAGCGGGGCTTAAGCATGTGCGGGGGAACAAG 6079  
 QY 1103 ----- 1103  
 Db 6080 GAGAGTGTGTTTCCAGGGCTGAGGGGTGCAAAAGTTCTCCGCTCCATGCGCTTGG 6139  
 QY 1103 ----- 1103  
 Db 6140 TTGCCCCCTCTCTCTTAATCAGTTATGAGCACTGTGTGTGAGTCTTGGGCTGCTTT 6199  
 QY 1103 ----- 1103  
 Db 6200 CTGCTGAGGGGAATATCCGCTCAAGAAATCATTAGGAGATATACAGTTACAAGAG 6259  
 QY 1103 ----- 1103  
 Db 6260 CTGTGTCACTGAGCATCTGTGTGCGGAGTCTGAAAACGAGAGAAAGTAGAGCC 6319  
 QY 1103 ----- 1103  
 Db 6320 TGAATCATTAATCTGGCAGGCTTCTGTGACAGTGAAGCTAGGGCAGTTAGAAGACT 6379  
 QY 1103 ----- 1103  
 Db 6380 CAGAGACTGTGGGTGTGTGTGGTGGAAAGAACAGCTGTGACAGCTTTGAGAC 6439  
 QY 1103 ----- 1103  
 Db 6440 GAGACAGAGAAACAGAGTGGGTGATGACAGACGGCTTGTGTGAGAGTAGTAG 6499  
 QY 1103 ----- 1103  
 Db 6500 AATTAAGATGATGAGGCTGAGGGTGAAGTGGCACTGGCTGACATGGCTGAGAT 6559  
 QY 1103 ----- 1103  
 Db 6560 GGGGCTGCTTTCAGGCTTCCGGCTTCAGTGAAGCAACGCTGTATCAGTACAGCCACC 6619  
 QY 1103 ----- 1103  
 Db 6620 GCAAGTAGTACACACAGATGCAAAATGTGTCCATGTCCCTTACTCCACCCC 6679  
 QY 1103 ----- 1103  
 Db 6680 ACCCAGAAACATGGGCTCTGCCCCGTACTGCTAGGTAAGCATTTGCTTTTATATC 6739  
 QY 1103 ----- 1103  
 Db 6740 TAGTTGGCTCCCTGCTATGATGTTTGTGTTGTTTGGACAGAGTCTCACTCTG 6799  
 QY 1103 ----- 1103  
 Db 6800 TCACCCAGGCTGAGTACAAATGGCGCATCTCGCTCACTGCAACCTTGCCTCGGGT 6859  
 QY 1103 ----- 1103  
 Db 6860 TCAAGCAATTTCTCTGCTCAGGCTCCTGAGTATCTGGACCTCTGATGATGATTT 6919



Db 9080 TTCAAGGGTACTACTGCGAGACCGAGTGGACCTCTGCTACTCGCGCCCTGTGGCCCC 9139  
 QY 1300 H1sg1yA3CyA8serA9glugl1ygl1yTtThCyA8y8sp01yTtThr 1319  
 Db 9140 CACGGGCGTCCGACGCGCGAGGGGGCTACACTGCTCTGTGTATGGCTTACAG 9199  
 QY 1319 ----- 1319  
 Db 9200 GGTGAGCCAAAGGAGGAGACTCATGGGCGAGCCCTGGAAAGGCTGACTGTGTGTGACAGC 9259  
 QY 1319 ----- 1319  
 Db 9260 ACAAATCAGAGCAAAATGCTGGCGGCTGCTCATTTCTTCCGAGTGAAGTGCAGTACA 9319  
 QY 1319 ----- 1319  
 Db 9320 CTGAGAGGTCAATAATCCGGCTGCTCCCTGACAGACCCCACTGAGAGGCAAGTCAA 9379  
 QY 1319 ----- 1319  
 Db 9380 GGAAGTCTTCTCCCTGCTGTGACAGAGACACAGGCAAGTGCATTGTGGGAGAGCTGG 9439  
 QY 1319 ----- 1319  
 Db 9440 CTCAGAGCAGTCCGCTGTGTGTCTGTATCAGCTTGGCCGCTAGTGTGTGACCT 9499  
 QY 1319 ----- 1319  
 Db 9500 TGAACAATTACTTAACCTTTCTGGCCCTCGGTTATTGAAAATGAGATATGACCCGTAC 9559  
 QY 1319 ----- 1319  
 Db 9560 TTAGGTCAATAGTGTGTGTGAGATTAATGTAATTAATTGATCTTGGAAACAGCTGG 9619  
 QY 1319 ----- 1319  
 Db 9620 TGGATGAGAAAGTGTGGCTTGAATTTAAGGAAACGCTGAGATGGGTGACTTGTCTGG 9679  
 QY 1319 ----- 1319  
 Db 9680 GTCTATCACAACCCGGTTTGAAGTGAAGGCTGGGCTGAGAACCTGTCTCTCCCACTG 9739  
 QY 1319 ----- 1319  
 Db 9740 AGCGAGTTCCTTTCTGAGCTGACAGAAATCTTTTCTCCGAGCATTAAGGGAGAAC 9799  
 QY 1319 ----- 1319  
 Db 9800 ACAGAAAGCTGGGGAACTGAGGGGGTTGAGTTCTGAGATGGGGTCTCCAGAAAAC 9859  
 QY 1319 ----- 1319  
 Db 9860 AATTGACCTGTATGGGCTTGGGCAAAATGGCTCCGCTCCGCGCAGCTGTGTAAAGA 9919  
 QY 1319 ----- 1319  
 Db 9920 GGATCTGGCTAGGGGCTTCCCTGTGGCTGTGACGCAAGGTCTTTACATCTGTCT 9979  
 QY 1319 ----- 1319  
 Db 9980 CCTGGAGTGTCACTGTCTGTCTGTATGTCTGACCCCTCCCTTGAAGCTCTGTGT 10039  
 QY 1319 ----- 1319  
 Db 10040 CTTCTCAGGACACCCCTGTCTCCCTTCCCTGTGGGCTTTCTTCTAGATGCCAAGC 10099  
 QY 1319 ----- 1319  
 Db 10100 CCTGCTCTTGTCTTCAAGGGCCATGCCCATGCTCTGACAGCATATAGGAGGCC 10159  
 QY 1319 ----- 1319  
 Db 10160 CACTCTTCCCGCTGCACTTCTGTGCCAAGATTGCCAAGCTTGACTTCTCTGTCTG 10219

QY 1319 ----- 1319  
 Db 10220 CTGTAGGGGTGGGGGAGAGGGAGAGACGGGCTTGGGGAGGGAGAGAAATCTTGTCTC 10279  
 QY 1319 ----- 1319  
 Db 10280 CAGAGAGATGGTGGGGGGGTTGGAGGGGGGGTTAGACAGCGCAACAAGTTTCC 10339  
 QY 1319 ----- 1319  
 Db 10340 ATGGAACAGAGCCAGTCACTAGGACAGAGCTGAGAACTTACTAGAAATCTGGGA 10399  
 QY 1319 ----- 1319  
 Db 10400 ATGGGAGAGAGAAAGGTGAATTTCCACTGANTGTGAAGTTTCACTCTCTGT 10459  
 QY 1319 ----- 1319  
 Db 10460 TGAATTCCTGTCTGTCTCAGGGCTCAGTTTCCCACTCTCAGCATGTCCAGGTGACAG 10519  
 QY 1319 ----- 1319  
 Db 10520 AGTCTTATGTGAGACCGCAGGGCTCAAGTGGAGTCTCTGCCATTCTGGGGGC 10579  
 QY 1319 ----- 1319  
 Db 10580 TCCCCAGTTCTCTCATTACTTTCCATTAACCAAGCAAAATCTAATGGAAGCTGTGTC 10639  
 QY 1319 ----- 1319  
 Db 10640 CCATCCCAAGGAGGAAACTGATCTGATGCTGAGAGGACAAAGACCAAGTGGAA 10699  
 QY 1319 ----- 1319  
 Db 10700 GAGGAGAGAGAGGCTCCAGCCCTCCGCTGTCTGACGGCCATCTGTGTGAC 10759  
 QY 1319 ----- 1319  
 Db 10760 ATTTAACCTTGGGGGCGCAGAGACACCCCAAGTGTCTGTGTGTGATGATGAAGGC 10819  
 QY 1319 ----- 1319  
 Db 10820 TCCGGATCCGGGCTTACAAACCAATGTCCAGAGGCTGTGCTTATGAACCCAAA 10879  
 QY 1319 ----- 1319  
 Db 10880 TGGGTAGGAAGGTGTGGGACTCCAGAGAACTTGCAGGGAGCAGTGTGTGAGCTG 10939  
 QY 1319 ----- 1319  
 Db 10940 TGGGAGAGCGGCTCAGTGAACCAAGGTGCCATGAAGGGCTGGGACATAGTTC 10999  
 QY 1319 ----- 1319  
 Db 11000 TCAGAGAAATGSCACACAGAGAGAGTGGGATGCTTGGAGGTTTGGGAGTATTACG 11059  
 QY 1319 ----- 1319  
 Db 11060 AGGTCACGGAGGCAATGGAGCTATCACTGGTGTGGGGGTCTCAAGTTCCCTCT 11119  
 QY 1320 ----- 1325  
 Db 11120 GTCTCAGTACTTGTACTCACTGTCTTCTCTTGTCTCAAGTGAAGCTGTGAGGTG 11179  
 QY 1326 Ser1Aa8sergl1yA9g1yThrProgl1yA1CyA1y8angl1ygl1yThrCyA8ValA8n 1345  
 Db 11180 AGTCTGTCTCAGGCGCTTGCACCCCGGTGTCTGCAAGATGGGGGCACTGTGTCAAC 11239  
 QY 1346 Leu1euVal1gl1ygl1yPhe1y8CyA8pCyA8Pro8er1yA8pPhe1y8pPro1yCy8 1365  
 Db 11240 CTGCTGGTGGCGGTTCAAGTGTGATGCCCATGTGAAGCTTGCAGAAAGCCCTACTCTG 11299

QY 1366 GlnVal1ThrTharSerPheProAlaHisSerPhe1LeThrPheArg1LeuArg1in 1385  
 Db 11300 CAGGTACACACGCGCAGCTTCCCCGCCACTCTTCATCACTTTGCGGCGCTGGCCAG 11359  
 QY 1386 ArgPheHisPheThrLeuAlaLeuSer----- 1394  
 Db 11360 CGTTCCACTTCACTCCGCGCCCTCTC- GTGAGTGGCTGGCACTGGGGGTGGGAGTGGG 11418  
 QY 1394 ----- 1394  
 Db 11419 CCGTGGGGCATCTGAGTGCCCTCTGTTCTGTGTGAAGTAAGATCAGGCACTGAA 11478  
 QY 1394 ----- 1394  
 Db 11479 AGCGTGAAGCAGGCTGCCGCGCAGAGCCAGGCTGTCTTGGAGTGCCTGTGTCTGCC 11538  
 QY 1394 ----- 1394  
 Db 11539 AGCAGATCTCCCTCCCTGCGCTGTGTAGGTGGCGGGGTGCTGGCATCCAGGTGGGTAC 11598  
 QY 1395 ----- PheAlaThrIysGluArgAspGly 1402  
 Db 11599 CCCATTCCCTGCCCCCATCCCACTCCTGTTCAAGTTTCCCAAAAGAGCGCCAGCGGG 11658  
 QY 1403 LeuLeuLeuTyrArg1YArgPheAsnGluLysHisAspPheVal1AlaLeuGluVal1Ile 1422  
 Db 11659 TTGCTGTGTACAAATGGGCGTTTCAATGAGAACATGACTTTGTGGCCCTCGAGGTGATC 11718  
 QY 1423 GlnGluGluVal1GluLeuThrPheSerAla----- 1432  
 Db 11719 CAGGAGAGGTCCAGCTCACTCTCTGCG- AGGTGATCAAGTGGCCCCCATCTTGGCC 11777  
 QY 1432 ----- 1432  
 Db 11778 CATCTTCAAAGGCCCAAGTCTTCCAGCCCTGAGCCCAAGCCACATACTATAGCC 11837  
 QY 1432 ----- 1432  
 Db 11838 AATCTGGGCGCAGCCCAAGCACTGGACCCCAAACTCTGCACTTCAAGCCCTGTCA 11897  
 QY 1433 ----- G1yGluSerThrThrValSerProPhe 1442  
 Db 11898 GGCATTCCAGCTCACTGCTGCTTCCCGGAGGAGTCAACCAAGGTGTCCTCATTC 11957  
 QY 1443 ValProG1yG1yValSerAspG1yG1nTyrPhe1SerVal1GluLeuLysTyrTyrAsnLys 1462  
 Db 11958 GTGCCGAGAGACTCAGTATGCGCAGTGGCATACGGTGAATAATCAATTAAG 12017  
 QY 1462 ----- 1462  
 Db 12018 GTGGGTGTGAGGAGCAGAGGTTGGGGTTCGTCTTGCCTCAGGTGTTACCCAGC 12077  
 QY 1462 ----- 1462  
 Db 12078 CCGTGAATGGCATTGCTCCAGGCTTGGGTGGGTGGGTCAAGGAGATTCTGGCTGAGAGCA 12137  
 QY 1462 ----- 1462  
 Db 12138 AGGAGACGGCTGGGAAGTGCACATTGCTGGGATTCCACCCCACTGGGCACTGTGC 12197  
 QY 1462 ----- 1462  
 Db 12198 ANGCTCTGTGTTCACTAGGACCGCAAACTTGAACACTGAGGAAGTATTAATGCA 12257  
 QY 1462 ----- 1462  
 Db 12258 TAAACACGTGAGCACAAGTGTGTATATGCAATGTGTGTGTAAGTGTGTGTGGCATGTGG 12317  
 QY 1462 ----- 1462  
 Db 12318 GTGCACACAGAGGACCAAGTGTGTGTGTGGCATGAACCTAGTGGGCTGTACTCAGTG 12377  
 QY 1463 -----ProLeuLeuG1yG1nThrG1yLeuProG1n 1472

Db 12378 TCCCCCTTCTGCTCTTCTCTGTCACAGCCACTGTGGGTCAAGAGGGGCTCCACAG 12437  
 QY 1473 G1yProSerG1yGluGluLysVal1AlaVal1Val1ThrVal1AspG1yCyAspThrG1yVal1Ala 1492  
 Db 12438 GGGCCATCAGAGCAGAAAGGTGGGTGTGTGACCGTGATGGCTGTGACACAGAGATGGCC 12497  
 QY 1493 LeuArgPheG1ySerVal1LeuG1yAsnTyrSerCyAspAla1AlaGlnG1yThrGlnG1yG1y 1512  
 Db 12498 TTGGCCTTCGATCTGTCTGTGGGCACTACTCTGTGTGCGGAGGCACCAGGGGTGGC 12557  
 QY 1513 SerIysLys----- 1515  
 Db 12558 AGCAAGAGTAGCAGGGGAAAGGCGAGGATGGGGTGAAGTAGAGGCAAGGCGACAGCA 12617  
 QY 1515 ----- 1515  
 Db 12618 GGAAGGAAGGTGAGAGGGGTGGGCTTTGGGGTCTCTCCCTTGTGCTCTTAGGCC 12677  
 QY 1515 ----- 1515  
 Db 12678 CTGCTCAAGCCCTGGGGATGGGGTCAAGACTGGGGAATCAGAGCAAGCCCTGTGGGGCAG 12737  
 QY 1515 ----- 1515  
 Db 12738 TCTGGGAGGGGGCTGAGTGTTCGACAAGAAAGATTCCAGGTAGGGGTGACGCCCTTCT 12797  
 QY 1515 ----- 1515  
 Db 12798 GAGCACTGCTGCAAGGGCTCTTGGACACTGCTACGCGGTGTGTGTCTGTGGCTGG 12857  
 QY 1515 ----- 1515  
 Db 12858 CGGACAGACGAGCGGAGGACCAAGAGCTCCAGGCCCTCCCTTGTGGCGCACCATCTTTG 12917  
 QY 1516 -----SerLeuAspLeuThrG1yProLeuLeuLeuG1yG1yVal1ProAspLeuPr 1532  
 Db 12918 CTCCCCAGGTCTCTGATCTGACCGGGGCCCTGTCTACGAGCGGGGTGCTGACCTTCC 12977  
 QY 1532 G1ySerPheProVal1ArgMetArgGlnPheVal1G1yCysMetArgAsnLeuGlnVal1As 1552  
 Db 12978 CGAGAGCTTCCCAAGTCCGAATGCCGCACTGTGGGTGTCATGCGGAACCTGTCAAGTGG 13037  
 QY 1552 PserArgHis1IleAspMetAlaAspPhe1AlaAsnAsnG1yThrVal1----- 1568  
 Db 13038 CAGCGGCAATAGACATGGCTGACTTCACTTGCACAAATGGCACCGTGCCTGTATGGG 13097  
 QY 1568 ----- 1568  
 Db 13098 GGCCTGGGTGAGCCAGGCTGGATGCCAGTGTGAGAGAGGGGCTGTGCTGTGTG 13157  
 QY 1568 ----- 1568  
 Db 13158 AGATATCCAGATGAGAGGTCTCAGAGGCTCTAAGTGTGCGCACAGGCTTCTCTTT 13217  
 QY 1569 -----ProG1yCyAspProAlaLysL 1575  
 Db 13218 AGAGCCCAAGCTCAGGCTCACTGAGCTGCTTCTCTCAATTCAGAGGCTGGCCCTGCAGAG 13277  
 QY 1575 YAsnVal1CyAspSerAsnThrCyHisAsnG1yG1yThrCyVal1AsnG1nTyrPAPA 1595  
 Db 13278 AGAAGTGTGTGACAGCAACTGTCACAAATGGGGGCACTTGTGCTGAACCAAGTGGAGCG 13337  
 QY 1595 LaPheSerCyG1yCyAspProLeuG1yPheG1yG1yLysSerCyAspAlaGln----- 1611  
 Db 13338 CGTTCAAGCTGGAGTGGCCCTTGGGCTTTGGGGGCAAGAGCTGGCCCAAGGGTGAAGGG 13397  
 QY 1611 ----- 1611  
 Db 13398 GCGGCTTTAGAGGCCACAGCTGGGTGCATCAACAGTCTGGGAATCGGAGGGTTGGG 13457  
 QY 1611 ----- 1611

Db 13458 GCAGGCACTGGGACAGGGCTCTGCTGAGCGGGCTGTGGGTGAAAAAGCGTGTCTGGGACG 13517  
 QY 1611 ----- 1611  
 Db 13518 AGCCCTGAAAGGGCAACACAGAGACAAAGGGCCAGTCAGAGGACGCTGGGCAACG 13577  
 QY 1611 ----- 1611  
 Db 13578 GTGGGGCGGTCCAGGGCAGGTACGCACTTTGAGAGGGCGGGCTGANTGAGGGAGTGGGC 13637  
 QY 1611 ----- 1611  
 Db 13638 TCTGCTCCCGTGCAGCGGCCACCGCTGAGCATCAAGCCCGGCTCTGAGGCTCC 13697  
 QY 1612 -----gluMetAlaasnProgluH1sPheleu 1620  
 Db 13698 ACCCGTCACAGTCTGCTTTTCTGCTCCCGAAGAAATGGCCAAATCCACAGACCTTCTG 13757  
 QY 1621 G1ySerSerLeuValAlaTrpH1sG1yLeuSerLeuPro11SerGlnProTrpTyrleu 1640  
 Db 13758 GGCACAGACCTGTGGCTGTGGCATGGGCTCTGCTGCCCATCTCCCAACCTGTATCCTC 13817  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspG1yValLeuLeuGlnAla11ethrArgG1y 1660  
 Db 13818 AGCTTCATGTTCCGACCGCGCCAGCGCCAGCGGTGTCTGTGCTGACAGCCATCAACAGGGGG 13877  
 QY 1661 ArgSerThr11eThrleu----- 1666  
 Db 13878 CGCAGACCATCACTCTACAGGTGATGCATGGAAGGGCGGCTGGCCCTGGCCATTA 13937  
 QY 1666 ----- 1666  
 Db 13938 GGGCCCTGTAGCTCTAGAGCGGCTGACAGAAAGTGGCTGGCAGAGTCCCTGGGACAGGGGT 13997  
 QY 1666 ----- 1666  
 Db 13998 GGGGACATATAGAGGCGCTGATCCGTTGGAGAGTCAATGCTGCACCTGTGGGCT 14057  
 QY 1666 ----- 1666  
 Db 14058 GAGGGAATTAATAGCTTCTCTCTATAGGCTCAGCTGGGACAGGGAGTGGTTGAT 14117  
 QY 1666 ----- 1666  
 Db 14118 GGTCCAGAAAGGGAACTGTGAGCTGAGAAACAGTGAATTAAAGCCGCTGCTGACCTGT 14177  
 QY 1667 -----GlnleuArgG1uG1yH1sValMetLeuSerValG1uG1yThr 1680  
 Db 14178 TTTCTTTCTCTCTGTGTGTAGGTTACGAGAGGGCCACGATGATCTGAGCGGTGAGGGCA 14237  
 QY 1681 G1yLeuGlnAlaSerSerLeuArgLeuGluProG1yAArgAlaAsnAspG1yAspTrpH1s 1700  
 Db 14238 GGGCTTCAGGCTCTCTCTCTCTGTGAGCGAGCCGGGCCAAATGACGATGACTGGCAC 14237  
 QY 1701 H1sAlaGlnLeuAlaLeuG1yAlaSerG1yG1yProG1yH1sAla11eLeuSerPheAsp 1720  
 Db 14238 CATGCAACAGCTGGCACTGGGAGCCAGCGGGGGGCCCGGCAATGCCATTCCTGTCCAT 14357  
 QY 1721 TyrG1yGlnGlnArgAlaG1uG1yAsnLeuG1yProArgLeuH1sG1yLeuH1sAlaSer 1740  
 Db 14358 TATGGGACAGCAAGAGGACAGGAGGCAACTGGGCCCCCGGCTGCATGTCTGCACTGGAGC 14417  
 QY 1741 Asn11eThrValG1yG1y11eProG1yProAlaG1yG1yValAlaArgG1yPheArgG1y 1760  
 Db 14418 AACATTAACAGTGGGGGAAATACCTGGGCGAGCCGGCGGTGTGGCCCTGTGGGGC 14477  
 QY 1761 Cys----- 1761  
 Db 14478 TGTTCGACAGTGAAGTCTGCTGCTGCTCCATCCCTCCCAACCACTGACAGCCC 14537  
 QY 1762 -----Leu-GlnG1yValArgValSerAspThrPr 1771  
 Db 14538 GCTCACTGAGGGCAACACTGCTCTGTCTCCCAAGGGTGTGGGAGGATACGCC 14597

QY 1771 OG1uG1yValAsnSerLeuAspProSerH1sG1yG1uSer11eAsnValG1uG1yCys 1791  
 Db 14598 GAAAGGGGTTAAACGCTCTGATATCCAGCATGGAGAGACATCAACGTGAGCAAGGCTG 14657  
 QY 1791 sSerLeuProAspProCysAspSerAsnProCysPro11aAsnSerTyrCysSerAsnAs 1811  
 Db 14658 TAGCTGCTGACCTGTGTGACTCAACCCGTTGCTGTCAACAGCTATTGCAAGCAAGCA 14717  
 QY 1811 PTrpAspSerTyrSerCysSerCysAsp----- 1820  
 Db 14718 CTGGACAGCTATTCTGCACTGTGA-TCCAGGTAATGTAAGATCCAGGGCAACGGGC 14776  
 QY 1820 ----- 1820  
 Db 14777 AGTTTACAGTGCCTGGGGCCACATGCTGCTGCTGTGTGACTTGGGCAATGGGCA 14836  
 QY 1820 ----- 1820  
 Db 14837 TCACACACACTCTGAGTCTTAGTTCCTACACAAAGCCATPAGCCAGCTGATAGA 14896  
 QY 1820 ----- 1820  
 Db 14897 TGTATGTGAAGAAACAATGAAGGTTGATGAAGATGTGCCAGGGAGGTGGG 14956  
 QY 1821 -----ProG1yTyrTyrG1yAspAsnCysThrAsnVa 1831  
 Db 14957 GGAATGAGCCTCTGCTGCTCTGCTGTGCTCCAGGTTACTTAATGTACACTTAAATGT 15016  
 QY 1831 1CysAspLeuAsnProCysGlnH1sGlnSerValCysThrArg1ySProSerAlaProH1 1851  
 Db 15017 GTGTGACCTGAACCCGTTGTGAGCACCAAGTCTGTGTATCCGCAAGCCCAAGTCCCCCA 15076  
 QY 1851 sG1yTyrThrCysGlnCysProProAsnTyrLeuG1yProTyrCysG1uThr----- 1868  
 Db 15077 TGGCTATACCTGCGAGTGTCCCAAAATACCTTGGGCCATATGTGAGACAGTAAAGC 15136  
 QY 1868 ----- 1868  
 Db 15137 AGACAGGGCATGTGGAGAGAGTCCAGTGGCTGTGCTCTCTGTGTGTGCCCTCA 15196  
 QY 1868 ----- 1868  
 Db 15197 GAGCCCGAAGCCTGCTGATCCACAGCGAGGTCAAGAGGCAATAGGGCTACCT 15256  
 QY 1869 -----Arg11eA 1871  
 Db 15257 AGTTAGGTGGAATGGCAAGGGAGCTCATGCTTACCTGGGTCTCTGTGCAAGATTG 15316  
 QY 1871 sPGLnProCysPProArgG1yTyrTrpG1yH1sProThrCysG1yProCysAsnCysAspV 1891  
 Db 15317 ACCAGCTTGTCCCGTGTGGCTGTGGGAGCATCCCAATGTGGCCATGCACTGTATG 15376  
 QY 1891 a1Ser1yG1yPheAspProAspProAsp1yAsn1yThrSerG1yG1yCysH1sCys1yS--- 1909  
 Db 15377 TCAGCAAAAGCTTGAACCAAGACTGCAACAAAGAACAGCGAGAGTGCACCTGCAAGGTGA 15436  
 QY 1909 ----- 1909  
 Db 15437 CAGCCCCAAGACAGCTTCCACTGTGGCCACTTGGGCTTGTCCACATCTCTGGGCCCT 15496  
 QY 1909 ----- 1909  
 Db 15497 AGAGGGCAGGCTTGAAGGATGGGTCTTCTGTAGAGCTGAGGGCTTGTGCTTGTCT 15556  
 QY 1909 ----- 1909  
 Db 15557 CACAGCAGGGCAGTTGGCTTACCTGTGTGTCTGAGGCGCAGGGGAGCCCTGGGC 15616  
 QY 1909 ----- 1909  
 Db 15617 CAATGTGAACCATTAAGCTCTGGCTGTGTGTGTGTGTGTGTGAGGCTGTCTGATTC 15676





```

Db 17837 TGAGAACGGGGCTTCTGGAATTCAGCCTGTGTCTTCTGGGCTCCAGCTGAGAGGCC 17896
Qy 2036 -----
Db 17897 GTCTCTACCTCTGAGGTCTTTTGTCCAGAGTTCTGTCCCTCACTCTGTCTCCG 17956
Qy 2037 -----
Db 17957 CTCGGCTGTGTAGCTGAGCGGCTACAGCGGAATGAGTCAAGGCTTACGCTCAGGGCCG 18016
Qy 2052 SerGlnGlnLeuAlaLeuLeuLeuArgAnaAlaThrGlnHisPheValIleGlyPheGly 2071
Db 18017 TCCACAGACTTACCCCTGCTCTCGCGCAACGCCAGCAGCAGCAGCTGCTTCTGCGC 18076
Qy 2072 SerAspValIleValAlaIleArgIleLeuAlaThrArgLeuLeuAlaHisGluSerThrGln 2091
Db 18077 AGGAGCTCAGAGTGGCTTACAGCTGGCCAGCGGCTGTCTGGCCCAAGAGCAGCCAG 18136
Qy 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlu----- 2106
Db 18137 CGGGGCTTGGGCTGTCTGTCCACACAGGACGTGCACTTCACTGAGGGTGGGCTTGGAGGA 18196
Qy 2106 -----
Db 18197 TGCAGGCTGCTGTGATTAGGGGTCAATGTAAGTGAAGCTTCAATGCAACCTGGG 18256
Qy 2106 -----
Db 18257 GGGCAGAGGGGGCCCTCCATCCACCTACAGAGAGCTCCCTGTGGCAGAGGCTCCA 18316
Qy 2106 -----
Db 18317 TGAACCCCTGTGACCCCTGCGCCAGCCCTTCCACCCCTTCTCATAGTAACATATGNG 18376
Qy 2106 -----
Db 18377 TGTGGGGGATGTCTGACCCACAGACAGAGCGCTGCTGTGTACCATGGGCTGGAGGA 18436
Qy 2106 -----
Db 18437 AGCAGATCTCAGACATGTGTGTGGGTTCAGGCAAGGGCTGGGAAGCTTATGTAGA 18496
Qy 2106 -----
Db 18497 GAATGAGAGAGGGCTTAGGGGCAAGACATCAGTGAACCTGATGGGCAGTGCGGTCCA 18556
Qy 2106 -----
Db 18557 GAGACATTTCAAGGGAGGCAAGTGAACAAGAGGAGGGGCTGTGGGGAGCAAGGA 18616
Qy 2106 -----
Db 18617 CTGGCCCGGGCAAGAGCGGGGCTCCATGTGGGGATGTCAAGTGTGGGTGTGAGACA 18676
Qy 2107 -----
Db 18677 CCCACCGTGAACCTTGCACCCACAGAACTCTGTGGGGGTGGCAGCGCTCTGAGAC 18736
Qy 2118 rAlaAsnIleValArgHisIleThrGlnLeuIleGlnGlnThrGlnIleValIleGlySerAlaLeuLeuAspThr 2118
Db 18737 AGCCACAGCGGCACTGGAGCTATTCAGGCAAGAGGGTGGCACCGCTGTGCTCT 18796
Qy 2138 uGlnHisIleArgIleValIleValIleAspValIleLeuAlaGlnAsnMetArgHisIleThrIleLeuSe 2158
Db 18797 CCAGCACTATGAGGCTTACGCGCAGTGCCTGGCCCAAGAACATGGGGCAACCTTACTTAAG 18856
Qy 2158 rProPheHisIleValIleThrProAsnIle----- 2167
Db 18857 CCCCTTACCAATGTCAAGCCCAACATTTGTAAAGCTGTGCTGTGGGTTGGGAGGGGTT 18916
Qy 2167 -----
Db 18917 TGTGAGAGGAGTCCCGACAGACGCGGTGTGTGGGGTCTGTGCTGCTCAAGGTCCG 18976

```

---

```

Qy 2168 -----
Db 18977 ATCTGTGACATCCCTTCTTATGATCTCGTGTGTGCTTGAACAAGGAACCTTGG 19036
Qy 2180 lAglyAlaIleLeuProArgIleArgIleAlaLeuArgIleGlnIleProProAspLeuGlnT 2200
Db 19037 CTGGGGCCAGGCTGCGCCGCTACAGAGGCCCTGTGGTGGGAGAGAGCCCGGAGCCTTGAGA 19096
Qy 2200 hTrThrValIleLeuProGluSerValIlePheArg----- 2210
Db 19097 CAACGATTTCTGTGCTGATGTGTCTTCAAG-AGGTCAGTGTGTGATGATGATGTTG 19155
Qy 2210 -----
Db 19156 GGAAGCTGAGACCCCAAGTGTGTGTGCAAGACTCCACAGAGAGAGGGGCCAGTTAAGTGTAC 19215
Qy 2211 -----
Db 19216 AGTGTCCCTCCAGAGAGAGCCCGCGTGTCAAGGCCCGAGAGGCCCGAGAGGCCAGG 19275
Qy 2226 lProGlnGlnLeuAlaIleArgIleArgIleAspGlnIleLeuSerGlnIleGlnIleValIle 2246
Db 19276 AGCCAGAGAGCTGGCAGCGGCAAGCGCACCCGAGCTGAGCCAGGGGTAGGCTG 19335
Qy 2246 aAlaSerValIleIleIleIleArgIleLeuAlaGlyIleLeuProHisAsnIleIleAspPro 2266
Db 19336 TGGCAGGCTCATCATCTACCGCAACCTGGCCGGCTACTGCTCTATTAATGACCTTG 19395
Qy 2266 sPlyAspSerLeu----- 2270
Db 19396 ACAAGCGCAGCTTGAAGGTCAAGCAGTGAAGGACAGAGTGTAGGGTATAGGTGCGGGC 19455
Qy 2270 -----
Db 19456 GGTGATGCTGAGGCAATGAGAGGGGTGCGGGGCGTCTCCCAATGATGATGCTGCGTGG 19515
Qy 2271 -----
Db 19516 TGACGTGCACTTGAAGTGGCCCAATGAGTGTGCTGCAAGTGTGCTGCAAGTGTGCTG 19575
Qy 2282 ValIleSerIleSerValHisAspAspGlnGlnIleLeuProArgAlaLeuAspIlePro 2301
Db 19576 GTGTGTGAGCATCAAGCTGCTCATGATGATGAGGAGCTTGTGCGCCGCGGCTGTGAACAACC 19635
Qy 2302 ValThrValGlnPheArgIleLeuGlnThrGlnIleValArgHisProIleCysValPhe 2321
Db 19636 GTCAAGGTGCAAGTGTGCGCTGTGTGAGACAGAGGAGGAGACCAAGCCATCTGTGTCTTC 19695
Qy 2322 TrpAsnHisSerIleLeu----- 2327
Db 19696 TGAACCATTTCAATCTGTGTAGAGCTGTGCTGCTGCGCCCTGAGGCTTGGGGCTGAAA 19755
Qy 2327 -----
Db 19756 GTCCAGGCCCCGTGACATCAACCTCTGTGTCTCCCTGACCCCTGACCTTCTCAACAG 19815
Qy 2328 ValSerGlnThrGlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 2347
Db 19816 GTCAAGTGCACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19875
Qy 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArg 2367
Db 19876 CAGCTCAAGCTGCGCAAGTGCACACATACAGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 19935
Qy 2368 ArgGlu----- 2369
Db 19936 CGGAGGTGTGGGCCCAAGAGGGGCACTGCAAGACCGGTGGTGGGCAACCAAGGCAAGGGG 19995
Qy 2370 -----
Db 19996 CTGGGTGCTCAAGTCTGCTTCTCTAATTCCTGTGGCCCTTGTGCACTTGTGAGAA 20055

```

QY	2370	GLGLIHLLEuLeuProLeuYstThLeuThrYValAlaLeuGlyValAlaThrlLeuAlaAl	2390
Db	20056	TGGGAGAGATCCCTGCACATGAAGACATGACATAGCGTGAAGTCACTTGGCTGC	20115
QY	2390	AlaLeuLeuLeuThrPhePheLeuThrlLeuLeuAlaGlyAlaLeuAyrSerAsnGlnHisG	2410
Db	20116	CCTTCTGTGCACCTTCTTCTTCTTCTCACTCTCTTGCGATCTGTGGCTCCAAACCAACCG	20175
QY	2410	YLIaAyrAyrAsuLeuThrlAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyLI	2430
Db	20176	CATCCGACGTAACTTGACATGCTGCCTTGGGCTGGGCTCACTGGTCTTCTCTGGGAAT	20235
QY	2430	eAsnGlnAlaAyrLeuPro-----	2436
Db	20236	CAACCAAGGCTGACCTCTCTGTAAATGCTCTTAATGCGCCAGAAACTGTCCCAACCTTCTC	20295
QY	2436	-----	2436
Db	20296	AGGCGGCTCCCGACGCCCCACGTGCAACCCCTGCTCTGCACACCATGAATCTTAATTAAG	20355
QY	2436	-----	2436
Db	20356	TGCCTAATGACAGCACTGGCCCCAGGGTTTCTCTTCTGTGGCTCCCCGGGAATCCCCAG	20415
QY	2436	-----	2436
Db	20416	CACCTGCCTTGAGCCCAAGGCTTCCCTGGAAGAGATTCCCAACCCAGAGCCCTCTTCAT	20475
QY	2436	-----	2436
Db	20476	GCCCTGACCCGAGACAGAGCTGTGCTGTGGCGGGCCCCGGTGCATGACTGCTGGCC	20535
QY	2437	-----PheAlaCyethrValIleAlaIleLeuLeuHisPheLeuThrLeuCyrt	2453
Db	20536	TGGGCGCCCTGAGTTGGCTGCACATTCATTCGTCATCTGCTGCACCTTCTTACTCTGCAC	20595
QY	2453	hrPheSerTrPAlaLeuLeuGlnAlaLeuHisLeuThrYrAlaAlaLeuThrGlnValAyrA	2473
Db	20596	CCYTTTCCCTGGGCTCTGCTGACGGCCCTTGACCTGTACCGGGCACTCACTGAAGTGGCGC	20655
QY	2473	srValaAnThGlyProMeArPheThrYrTrMetLeuGlyTrPGLYValProAlaPheI	2493
Db	20656	ATGTCACACCGGCGCCCATGGCTTCTACACATGCTGGGCTGGGGGCTGCTGCTTCA	20715
QY	2493	lethr-----	2494
Db	20716	TCACAGGTACTCCACCCCATTCGCCAGTCTTGAGGTCCACATCCCTGAGTCCACTTTGT	20775
QY	2494	-----	2494
Db	20776	GCCATGTTCTTCCACCCACATPACAGGCGCTGAGGCCCACTCCATGCCCCOAGCGC	20835
QY	2494	-----	2494
Db	20836	CCTTAATTCAGAGTGTCCCTGTGGTTAAACCCAGACTCTGAGCGCGCACCCAGGCGTCA	20895
QY	2494	-----	2494
Db	20896	CTCCCTTAATCTGGGAGAGATCGTAGGGCCGATGTGGGCGAACAACCTTTTCCATGCTT	20955
QY	2494	-----	2494
Db	20956	CATGCGTGGCCCTGTGAGCCCACTGGCCGAGGCCCTACTTCCAGGCCCTCTCATCAACC	21015
QY	2495	-----GlyLeuAlaValGlyLeuAspProGluG	2504
Db	21016	CTCCACATGCTCCGCTGTCTTCATGCTCAAGGCGTA-GCGGTGGGCTGTGAACCCGAGG	21075
QY	2504	LYTYrGLYAsnProAspPheCyrtTrpleuSerLleYrAspThrLeuIleTrpSerPheA	2524
Db	21075	GCTAGGGGAACCTGACTTCTGCTGGCTCTTCATCTATGACAGCGTCACTGGAGTTTG	21135
QY	2524	laGlyProValAlaPheAlaValSer-----	2532

Db	21135	CTGGCCCGGAGGACTTTTGGCGTCTCGGTAGAGTCTAGCAGAGTGGGTGTCACTGT	21194
QY	2532	-----	2532
Db	21195	GGCCCTCCTTGTGTCGTCTCTGTCTGTGGGCGCGGTGTCTCAGAGACTCCCTTAGGA	21254
QY	2532	-----	2532
Db	21255	ACAGGTGAGGCACAGGGGCGCTGTGGGCTGGGTGAAGCTGTTTGTCCCACTAGCACCC	21314
QY	2533	-----MetSerValPheLeuTyrIleLeuAlaIaArgAs	2545
Db	21315	CCATGCCCTCTGCCCGCTCGCTAATAGATGTCTTCTGTACATCTGTGGGCGCCGGCCT	21373
QY	2545	erCysAlaIaIaGlnArgGlnGlyPheGluValysGly-----	2557
Db	21375	CTGTGTGTGCCACAGGCGAGGGCTTTGAGAAAGAAAGTCTGTGTGATATAGGTTGGG	21436
QY	2557	-----	2557
Db	21435	GTGCTGTGGCCATGGGCGAGGCACAGCATGGAGGTCTCATGGCCGAGACTATGGCTGT	21494
QY	2558	-----ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT	2575
Db	21495	CCCTATCCCAAGC-TGGGGCTGTGACCCCTCTTGCCGCGTCTCTGTGTGTGAGGCCA	21553
QY	2575	hrrTyrLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheA	2595
Db	21554	CGTGGCTGTGGCACTGTCTCTGTGTCAACAGCGACACCTCTCTTCCACTACCTCTTGG	21613
QY	2595	IaThrCysAsnCysIle-----	2600
Db	21614	CTACTGTCAATTGTATTCATTCAGATACCTTGCCCAAGCTGTGGAGAGGAGGACCTGGGCT	21673
QY	2601	-----GlnGlyProPheIleP	2606
Db	21674	GTGGATGCTCGAATATATGACACAGACCGTGTGTCTTGTGCTGTGCGACAGGCGCCCTTATCT	21733
QY	2606	heLeuSerTyrValIleLeuSerLysGluValArgIleValLeuIleValLeuAlaCysSerA	2626
Db	21734	TCCCTCTCTATGTGTGTGTGTATGAGAGAGAGGTCCGGAAGACACTCAGACTGTGCGTGCAGCC	21793
QY	2626	rdLysProSerProAspProAlaLeuThrThrLysSerThrLeuThrSer-----	2642
Db	21794	GCAAGCCACGCTTGAACCTGTCTGTACACCAAGTTCACCTGTACCTCGGTGAGGGAAC	21855
QY	2642	-----	2642
Db	21854	CAGGGGTCTCAGAGAGGGCGGTGAAGGAGGAGAGGCAAGAGGCCAGTGAAGCTGTACT	21913
QY	2642	-----	2642
Db	21914	TTTGGCCCACTCCCTTCTCTTTTCTGCATCCCTTCTGTGAAGGTGAGAGGGAGGTG	21973
QY	2642	-----	2642
Db	21974	AAATGCTGTGTGGGCTGTGGGAGAGAGATGTGAAGAGGATGTATAGATTGTGACTTCA	22033
QY	2642	-----	2642
Db	22034	GGAGATGGGGCTGAGACAGATATATCAGGTATATATAGTGCCTTGCACTGAGATCTCTCA	22093
QY	2642	-----	2642
Db	22094	TGAATTTTATGTAATTTTCTGTCTCCCAAGGTGTCTCTCAGAGACAATTTCTGTGTCTAT	22155
QY	2642	-----	2642
Db	22154	CTTTTACAGAGCAATTATATATAGTATCCAGATCTCTGGGGCATGACGGGGTAGACA	22213
QY	2642	-----	2642



<http://www.sanger.ac.uk/HGP/Chr1>  
RP11-29704 is from the library RPC1-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6  
This sequence is the entire insert of clone RP11-29704. The true  
left end of clone RP11-173K24 is at 98796 in this sequence. The  
true right end of clone RP11-352P4 is at 101180 in this sequence.  
location/Qualifiers

## FEATURES

```
source
1..169241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-29704"
/clone_1lb="RPC1-11.2"
1676..11775
/note="MIR repeat: matches 45..152 of consensus"
misc_feature
2574..3410
/note="CpG island"
/evidence=not_experimental
2809..2936
/note="64 copies 2 mer gg 60% conserved"
repeat_region
2814..2943
/note="13 copies 10 mer ggcggggcg 63% conserved"
2982..2987
/note="Sequence from overlapping clones BA52P4 (
AL556389) and d641D2 (AL157901). Assembly confirmed by
restriction digest."
3007..3116
/note="11 copies 10 mer gccggggcg 63% conserved"
3364..3411
/note="16 copies 3 mer tgc 83% conserved"
9894..10207
/note="AluSg repeat: matches 1..296 of consensus"
10332..10617
/note="Aluo repeat: matches 10..299 of consensus"
11683..12328
/note="CpG island"
/evidence=not_experimental
12551..12727
/note="MIR repeat: matches 10..197 of consensus"
13355..13442
/note="44 copies 2 mer gg 65% conserved"
15405..15448
/note="22 copies 2 mer gt 75% conserved"
17917..17995
/note="MIR repeat: matches 45..129 of consensus"
19254..19293
/note="4 copies 10 mer agcctggcg 95% conserved"
19651..19956
/note="AluSg repeat: matches 5..309 of consensus"
19972..20270
/note="AluSc repeat: matches 1..296 of consensus"
26166..26195
/note="10 copies 3 mer gag 90% conserved"
29612..29726
/note="FLAM C repeat: matches 1..118 of consensus"
29778..30077
/note="AluSx repeat: matches 1..298 of consensus"
30135..30428
/note="AluY repeat: matches 1..298 of consensus"
30454..30616
/note="MIR repeat: matches 98..256 of consensus"
32159..32470
/note="AluSd repeat: matches 1..312 of consensus"
32481..32780
/note="AluSx repeat: matches 1..298 of consensus"
34670..34803
/note="MIR repeat: matches 123..250 of consensus"
36339..36835
/note="CpG island"
/evidence=not_experimental
```

```
repeat_region
37313..37937
/note="MER41B repeat: matches 1..635 of consensus"
repeat_region
37938..38051
/note="MER41B repeat: matches 103..213 of consensus"
repeat_region
38052..38351
/note="AluY repeat: matches 1..300 of consensus"
repeat_region
38352..38753
/note="MER41B repeat: matches 213..635 of consensus"
repeat_region
38987..39182
/note="MER41B repeat: matches 1..171 of consensus"
repeat_region
39183..39482
/note="AluSx repeat: matches 1..300 of consensus"
repeat_region
39483..39504
/note="MER41B repeat: matches 170..191 of consensus"
repeat_region
39537..39824
/note="AluSx repeat: matches 1..304 of consensus"
repeat_region
39850..40383
/note="MER41A repeat: matches 1..554 of consensus"
repeat_region
40394..40483
/note="MER41A repeat: matches 359..448 of consensus"
repeat_region
40532..40838
/note="AluY repeat: matches 1..311 of consensus"
repeat_region
40956..41169
/note="L1P1 repeat: matches 5927..6151 of consensus"
repeat_region
41170..41472
/note="MER41D repeat: matches 433..745 of consensus"
repeat_region
41473..41771
/note="AluY repeat: matches 2..300 of consensus"
repeat_region
41772..42011
/note="MER41D repeat: matches 745..981 of consensus"
repeat_region
42399..42702
/note="AluSx repeat: matches 1..304 of consensus"
repeat_region
42948..43256
/note="AluY repeat: matches 1..310 of consensus"
repeat_region
44024..44141
/note="MIR repeat: matches 21..147 of consensus"
repeat_region
44187..44622
/note="MIR repeat: matches 77..547 of consensus"
repeat_region
44790..45097
/note="AluSx repeat: matches 1..306 of consensus"
repeat_region
45098..45232
/note="AluSg repeat: matches 1..136 of consensus"
repeat_region
46712..47013
/note="AluY repeat: matches 1..305 of consensus"
repeat_region
47032..47323
/note="AluSx repeat: matches 1..292 of consensus"
repeat_region
47659..47826
/note="L1 repeat: matches 2545..2709 of consensus"
repeat_region
48017..48204
/note="MIR repeat: matches 22..209 of consensus"
repeat_region
49692..49988
/note="AluSx repeat: matches 1..297 of consensus"
repeat_region
54709..55019
/note="AluSg repeat: matches 1..311 of consensus"
repeat_region
55261..55380
/note="60 copies 2 mer gt 79% conserved"
repeat_region
55267..55386
/note="12 copies 10 mer gtggtgtgtg 79% conserved"
repeat_region
55386..55419
/note="L1 copies 2 mer gt 88% conserved"
repeat_region
55756..56029
/note="L1 repeat: matches 46..328 of consensus"
repeat_region
56258..56327
/note="MIR repeat: matches 323..388 of consensus"
repeat_region
56425..56915
/note="L1 repeat: matches 330..834 of consensus"
repeat_region
56926..57541
/note="L1 repeat: matches 2097..2748 of consensus"
repeat_region
58447..58523
/note="MIR repeat: matches 49..146 of consensus"
repeat_region
59152..59300
/note="MIR repeat: matches 29..185 of consensus"
repeat_region
60773..61014
```

```

repeat_region /note="MIR repeat: matches 4. .262 of consensus"
61115. .61227
repeat_region /note="MIR repeat: matches 36. .150 of consensus"
61443. .61519
repeat_region /note="MIR repeat: matches 6234. .6314 of consensus"
61956. .62063
repeat_region /note="MIR repeat: matches 35. .141 of consensus"
63402. .63437
repeat_region /note="18 copies 2 mer ta 80% conserved"
68939. .69310
repeat_region /note="AluB repeat: matches 2. .306 of consensus"
69476. .69781
repeat_region /note="AluX repeat: matches 1. .292 of consensus"
70577. .70718
repeat_region /note="MIR3 repeat: matches 2. .155 of consensus"
71406. .71699

Alignment Scores:
Pred. No.: 0 Length: 169241
Score: 12339.00 Matches: 2907
Percent Similarity: 36.92% Conservative: 0
Best Local Similarity: 36.92% Mismatches: 8
Query Match: 9 Gaps: 4967
DB: 32

US-09-916-849a-3 (1-2923) x AL390252 (1-169241)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20
DB 3330 ATGGGAGAGCCGGGCGACCGGGGCTCCCTCCCAAGCCGCGCCGCTCTGCTG 3389
QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSer 40
DB 3390 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449
QY 41 GlySerArgGlyArgGlySerSerGlyValCysAlaProMetGlyTyrProCysProSer 60
DB 3450 GGGTTCAGAGGAGCGAGGCTCTTGGGGGCTGCGCCCGCATGGGCTGCTGCTGCT 3509
QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80
DB 3510 TCAGGCTGGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3569
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGlnSerGlnAlaHis 100
DB 3570 GGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3629
QY 101 IleProLeuProProAlaProGlnGlyCysProTyrProCysArgLeuLeuGlyTlGly 120
DB 3630 ATTCCTTACCAACGAGCTCTGAGAGCTGCTGCTGAGCTGCTGCTGCTGCTGCTG 3689
QY 121 GlyHisLeuSerProGlnGlyLeuLeuThrLeuProGlnHisProCysLeuValAla 140
DB 3690 GGCACCTTCTCCCAAGGAGGAGCTCACTGCTGAGAGCACTGCTGCTGCTGCTG 3749
QY 141 ProArgLeuValGlyGlnSerCysLeuLeuAlaGlnAlaProGlyLeuArgAlaGlyGln 160
DB 3750 CCAGGCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3809
QY 161 ArgSerProGlnGlnSerLeuGlnGlyValArgArgLeuAsnValLeuThrAlaProGln 180
DB 3810 AGGTTCACCAAGAGTCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3869
QY 181 PheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200
DB 3870 TTCAGAGCCCGCAGCTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3929
QY 201 AlaSerLeuValAlaIleAspProAspGlnGlyGlnAlaGlyArgLeuGlnTyrThrMet 220
DB 3930 GCATCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3989
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240
DB 3990 GATGCCCTCTTTGATGACCGCTCCACAGTCTTCTCCCTGAGCCAGTCACTGTGCA 4049
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyThrLeuSerThrHisValPheArgValThr 260
DB 4050 GTAAACCAAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4109
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280
DB 4110 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4169
QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlnGlnGlnGlnGlnGln 300
DB 4170 GACCAATGACCAATGACCTGTGTGTGACGACGACGACGACGACGACGACGACG 4229
QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320
DB 4230 AACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4289
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGlnValPhe 340
DB 4290 AATGCCAATATTTCTGACCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4349
QY 341 GlnIleAspProAspSerArgValIleArgThrArgGlyProValAspArgGlnGlnVal 360
DB 4350 GAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4409
QY 361 GluSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyValArgAspProGlyProArgSer 380
DB 4410 GAATTCACCAAGCTGACGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4469
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAspAspAspAspAlaProGlnPheSer 400
DB 4470 ACCAAGCCCGCTGCTTCTTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4529
QY 401 GlnIleAspArgTyrValGlnValArgGlnValAspValThrProGlyAlaProValLeuArg 420
DB 4530 GAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4589
QY 421 ValThrAlaSerAspAspAspAspAspGlySerAsnAlaValAlaHisTyrSerIleMetSer 440
DB 4590 GTCAAGCTCTCGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4649
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValAlaSer 460
DB 4650 GGCATATCTGCGGAGACGTTTATGATGATGATGATGATGATGATGATGATGATG 4709
QY 461 ProLeuAspTyrGlnThrThrThrThrThrThrThrThrThrThrThrThrThrThr 480
DB 4710 CTTCTGACTATGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 4769
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
DB 4770 GGTCCCACTCTCTATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4829
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520
DB 4830 AATGCCCCCATCTTCTGACGACCCCTTTCAGGCTACGTCCTGAGAGAGTGTCCCT 4889
QY 521 GlyTyrLeuValLeuHisValAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540
DB 4890 GGCTACCTGATTCATGATTCAGGCTGATTCAGGCTGATTCAGGCTGATTCAGGCTG 4949
QY 541 GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGly 560
DB 4950 GAATACCGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5009
QY 561 TrpIleSerValAlaAlaGlnLeuAspAspGlnGlnValAspPheTyrSerPheGlyVal 580
DB 5010 TGGATCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5069
QY 581 GlnAlaArgAspHisGlyTyrThrProAlaLeuThrAlaSerAlaSerValThrVal 600
DB 5070 GAGGCTGAGACCATGACCTCAAGCACTCACTGCTCGGCAAGTGTCAAGCTGATG 5129

```

QY 601 LeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 Db 5130 CTGGAGTGTCAAGCAACATCCACCTTTACCCACAGAGTACCACTGGGGCTGAT 5189  
 QY 621 GluAspAlaValAlaGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 Db 5190 GAGGAGTCAAGCTGTGGCACACAGCTGTGTGACCGTGTGACCGTGTGACCGTGTGAT 5249  
 QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 5250 AGGTGATACCTTACCAAGATCACAGTGGCAATCTCGAACCCTTCTTCATCACACAG 5309  
 QY 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleuGlnIleuArgGln 680  
 Db 5310 CAAGAGTGTGTGGCTGGTATCCCTTGGCTCCACTGGAACAACTTGAGCGGGCAG 5369  
 QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 5370 TATGTGTGGCTGTACCGCTCCGATGGCACTGGCAGGACACGGCACAGATTTGTGTG 5429  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 Db 5430 AATGTACCGAGCCCAACACCCATCGTCTCTTCAAGCTCCCACTATACAGTGAAT 5489  
 QY 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db 5490 GTTATAGAGGACCGGCGGGCAGGACACAGGTGTGTCTGATCAGCCGACGAGTGAAGAC 5549  
 QY 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheAlaGly 760  
 Db 5550 ACGGGTAGAAGATCCCGCATCACTACTCATGAGAGACAGCATCCCGCATGCCCATC 5609  
 QY 761 AspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db 5610 GATGACGACACGGGGGCTGTACACCCAGCTGAGCTGACATATGAAGACCAAGTGTCT 5669  
 QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
 Db 5670 TACACCTGGCCCATCTGTCTGGGACAAATGCGATTCGCCAGAGTCCGACACACTAC 5729  
 QY 801 LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db 5730 CTGGAAATCTCGTGTGAACGAGTGAATGACAAATGCCCTTCAGTCTCTGCGAGCTCTTAC 5789  
 QY 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 5790 CAGGGCAGTGTCTATGAGATGTGCCACCTTCACTAGCGTCTTGCAGATCTCAGCCACT 5849  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 5850 GATCGGATTTCTGACCTTAATGGCAGGGTCTTCAACCTTCCAGAGAGGCGACGATGGA 5909  
 QY 861 AspGlyAspPheIleValGlnIleThrSerGlyIleValArgThrLeuArgAspGlyLeuAsp 880  
 Db 5910 GACGGTGACTTATGTGTGAGTCCAGTCCAGGATGTGTGAAACGCTACAGAGAGCTGTGAT 5969  
 QY 881 ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspIleGlyMetProPro 900  
 Db 5970 CGAGAGACCTGGCCCGATATGTCTTGGCGGCAATATGCACTGACAAAGGAGAGGCCCA 6029  
 QY 901 AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920  
 Db 6030 GCCCGCACACTATGGAATGACAGTCACTGTGTGATGTGAATGACAAATCCCGCTGTC 6089  
 QY 921 PheGlnGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940  
 Db 6090 TTGTAGAGGATATGATTTGTGTGTGTGTGAAGAGAACGCCCATTTGGAGCTGAGCGTG 6149  
 QY 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960  
 Db 6150 GCCCGGATCAAGACACTGACCCCGATGAAGGACCAATGCCAGATTATATACAGATT 6209

QY 961 ValGlnGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980  
 Db 6210 GTGAGGGGCAACATCCCTGAGGTCTTTCAGCTGGACATCTTCCGGGGAGCTGACAGCC 6269  
 QY 981 LeuValIleAspLeuAspTyrGluAspArgProGluTyrValIleValIleGlnAlaThrSer 1000  
 Db 6270 CTGTGATGACTTATGACTTACAGAGACCGGGCTGAGTACGTGTGTATCTCAGGCCATGCA 6329  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 6330 GCTCTGTGTGAGCGGGCTTACAGTCACTCCGCCCTCTTTCAGCGGAATGACAAACCA 6389  
 QY 1021 ProValLeuGlyAsnAspGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1040  
 Db 6390 CCAAGTGTGGGCAACTTGTAGATCTTTCAACAACATATGACCAATCGCTCAAGAGAC 6449  
 QY 1041 PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
 Db 6450 TTCCCTGGAGGTCCTATTTGGCTGGAGTACCTGCCATGACCTGATATCTCAGATATGTCG 6509  
 QY 1061 ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValIleLeuAsnAlaSerThrGly 1080  
 Db 6510 ACTTACAGCTTTGAGCGGGGAAATGAACTCAGCTGTCTGTCTCAATGCTCCACGGGT 6569  
 QY 1081 GluLeuIleuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
 Db 6570 GAGCTGAAGCTTAAAGCCGCGCACTGGACAAACCGGCTCTGAGGCGCATATGAGCGTG 6629  
 QY 1101 LeuValSer----- 1103  
 Db 6630 CTGTGTCT-AGGTAAAGAAAGGAGCCAGGTGGCGCTGGGTGGGTAGCTGCGGGGATG 6688  
 QY 1103 ----- 1103  
 Db 6689 GTCTGGGAGGACATGAGGATAGGGTGGATCCAGAAAGACATACAGATCCACTCCT 6748  
 QY 1103 ----- 1103  
 Db 6749 GCCCAGTCTGGCAGAGAGAGAGGGCAAGACCGTTGGAGAGAGCCCGCAGAGAA 6808  
 QY 1103 ----- 1103  
 Db 6809 CCTGGCGGCTGTGTGAGGCGCCCACTTCTGTGTGGCTGTGGCTGTGCTTCCGTTGGC 6868  
 QY 1103 ----- 1103  
 Db 6869 TGCCCCCACTGTCTACAGCGCTGTGAGGTGTGGGAGAAAGTGTGTGAGGCTGCCGCG 6928  
 QY 1103 ----- 1103  
 Db 6929 TATGCTGCAAGATTGTGAAAAAGCGCGTGTGAGCTGTGGGGGCGCAGAGAGCTGG 6988  
 QY 1103 ----- 1103  
 Db 6989 GGGAAAGGCAAGAGAGACTGTGTGCGCAGAGAGAGAGGCGAGGCGCCCTTACACCC 7048  
 QY 1103 ----- 1103  
 Db 7049 CATCAAGATATTCTGTGAGTCTACACAGACCTTTGGCAGAGCTGTGGAGCAACGGGAGAC 7108  
 QY 1103 ----- 1103  
 Db 7109 AGGGCTGCGCACCCAGGGGTCAAGAACCTGTATGCTCTTCTGAGCCCTTTTGTCTTC 7168  
 QY 1103 ----- 1103  
 Db 7169 TGGCTGAGCTGCAAGATGGGCTGACAGGAGACTTTGTCTTATGAGCATTTCTTTCCCTCAG 7228  
 QY 1103 ----- 1103  
 Db 7229 CAGAAAGAGCTGTGCTGGCTGCCCTTCAAGAGCTCCACAGAGAGTGAAGGCAAGAGCTC 7288  
 QY 1103 ----- 1103

Db 7289 ATTGTCCTCAACAGACCCTTTCAGAGTCCTGGGTTGGGCTTGCTTGAGAG 7348  
QY 1103 ----- 1103  
Db 7349 GGGCTGATGAAGCTGGGGGAGAGCTGAGGCTGTGGGCCCCCTTGCCACTCATG 7408  
QY 1103 ----- 1103  
Db 7409 GCGTGTCTTAGGGTGCAGGCTTCTTCTGCTAGCTGTGACCAAGCCAAAGCTCCAG 7468  
QY 1103 ----- 1103  
Db 7469 GAATGAGTGTAGGAGCAGAGACGGCCAGTGACTGTGACTTTGAGGAGAGAGACC 7528  
QY 1103 ----- 1103  
Db 7529 CAAGGACCTTATCTGAGAGAGAACAGAGTCCAGGGGCTGGAGCTGGCCGAACCC 7588  
QY 1103 ----- 1103  
Db 7589 AGGCTGGGGCAATTCCTGCTGCTGTGGCCAGCACTCAGCTTACAGACTCCCTGAGACAG 7648  
QY 1103 ----- 1103  
Db 7649 GGCTGGGGGAGAGAGTTGAGAGGGGTTGGTGAAGATTTCTGTCTTTTATCAGCTG 7708  
QY 1103 ----- 1103  
Db 7709 TGGGAGCCAGAGTACAGAGGGAGAGAACCAACCAAGTTGCTGAGTGCCTTCCC 7768  
QY 1103 ----- 1103  
Db 7769 AAGCTGGAGCACTGAGACCCCTCATGAGGGAGTGTGTAGTGCACCCAGAGGGTGG 7828  
QY 1103 ----- 1103  
Db 7829 ATCAGTCAATCCCTCTAGGAGCTTCAATCCATCAGGAGAGGAGAGAGAGAGAG 7888  
QY 1103 ----- 1103  
Db 7889 GGCTGGAGGGCCCTTACTCCACCCAGGAGAGTGTGGCTTGAAGGGAGAGAGGGTATG 7948  
QY 1103 ----- 1103  
Db 7949 CCAGAGTCAGGGCTGCTCACTCAGGGTCTTGTGTCGACAGAGCTCCTGCCCCAC 8008  
QY 1103 ----- 1103  
Db 8009 CCTTCAAGTACCCACAGAGAACAGAGAGCCTTTTCAAGGACTGTGGCTGAGGTT 8068  
QY 1103 ----- 1103  
Db 8069 CTGAGGCTCTGAGACTTCAAGTTTCCCTCAAGAGGGCGAGAGAGGCTGAGCTCC 8128  
QY 1103 ----- 1103  
Db 8129 TGTGTCTCACTGTGCTGAGCACTTGAGCAAGTGGAGAGAGTGGAGAGAGGGCC 8188  
QY 1103 ----- 1103  
Db 8189 TCTCCCTCTGTGCTTCACACCAAGCTGCTTTGTGTAGCGTGGCCAGGCACTGGCAGA 8248  
QY 1103 ----- 1103  
Db 8249 GAAGGCCCCCATTAAGCTTGGGTCTCCACAGGGAGAGCCCTTGAAGCTTGGTTCC 8308  
QY 1103 ----- 1103  
Db 8309 TAGTCTAGAGAGGCTCGGGGACAATCCAGAGACGGCTCTCTGACCAGGAGCTCTGG 8368  
QY 1103 ----- 1103

Db 8369 CCCACCTCTCCCTTGAGCCCTGACCTTGAAGCACTCAACAGCAACCCCAAGTT 8428  
QY 1103 ----- 1103  
Db 8429 CAGGAGCCGGGGTGAAGGGAGCTGACTCCACTGCCCTTTGTTTCTCTCTTGT 8488  
QY 1103 ----- 1103  
Db 8489 TCTTGTCTTCTTCTTCTTCCCGCCCTGCCATGAGGAGCCCTGACACTGCTCTC 8548  
QY 1103 ----- 1103  
Db 8549 TCCGGTCCCTCTCCCTGCTCTGAGGCTCTCCGCTCTCCACTTCCACTTGTCTGCT 8608  
QY 1103 ----- 1103  
Db 8609 CCTCTCTGAGGCTTGTGCTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8668  
QY 1103 ----- 1103  
Db 8669 CCTATGCTGAGCTGCCAGCCTGGGCGATGGCTCCCTGAGGCTGAATTCCTCTC 8728  
QY 1103 ----- 1103  
Db 8729 GCTCTTTCTTCTCTGCTCAGACACGACAGAGCTCTGTTCATGTCTGGGACAGCTCC 8788  
QY 1103 ----- 1103  
Db 8789 AGAACAGAGAGCCCTTGGCTTCAGTCTCCGCTGACATCTCCAGGGAGAGAG 8848  
QY 1103 ----- 1103  
Db 8849 GCTGGAGCTGACCGGAGCTGACATTTTCAAGTGTCTGGCTTAACTTGGACAC 8908  
QY 1103 ----- 1103  
Db 8909 AGTGGCATTTCTTGGGCTCAGAGGCCATCCAGCTGTTCTTTTCTGTCCAGAGACA 8968  
QY 1103 ----- 1103  
Db 8969 GCTCAGCAAGTACTCCCAACCCCTGTTCCTTCATCTGCTTGTGACCTGAGGGGTGAC 9028  
QY 1103 ----- 1103  
Db 9029 TAGTCAAGCAGGGGAGGTGATATAGAGAGAGCAGTGGGTTGACTGTGGGA 9088  
QY 1103 ----- 1103  
Db 9089 TGGGAGACACAGACAGAGGGGCAGTGGCCACAGATGTGACCAATGGAGAGACA 9148  
QY 1103 ----- 1103  
Db 9149 GAGCATGTGTGGAGAGGGGCCAGGGGCTTAAGACATGGCGGTGGGGACACAG 9208  
QY 1103 ----- 1103  
Db 9209 GAGAGTGTGTTCCAGGCTGAGGGGTGGCAAGTTCTCCCGCTCATGSCCTTG 9268  
QY 1103 ----- 1103  
Db 9269 TTGCCCCCTCTCTCTCTTAATCAATTATGGGCACTGTGTGTGAGTCTTGGGCTGCTT 9328  
QY 1103 ----- 1103  
Db 9329 CTGCTGAGGGAATATCCGTCAAAGAAATATTAGGAGATATATACAGGTATACAGAG 9388  
QY 1103 ----- 1103  
Db 9389 CTGTGTCAAGTGAATCTGTGTGGGGAGTCTGAACGAGAGAAAGTAGACC 9448  
QY 1103 ----- 1103  
Db 9449 TGATGACATAAATCTGGGCAAGCTTCTTGAGACAGTGTGAGTGAAGGCAAGTTAAGACT 9508



QY 1103 ----- 1103  
Db 9509 CAGAGACTGTGGCTGTGTCTGTGGTGGAGGAAACGCTGTGCAGCAGCTTTAGAC 9568  
QY 1103 ----- 1103  
Db 9569 GAGAGCAGAGAAACAGAGTGGGTATGACAGCAGGCTTGTGTGAGAGCTAGTAG 9628  
QY 1103 ----- 1103  
Db 9629 AATTAAGCAATGATGAGCTGAGGGTAGGTGGCAGTCCGCTGAGGACATGCTGAGAT 9688  
QY 1103 ----- 1103  
Db 9689 GGGGCTGTCTTCAAGCTTCCGGCTTCAAGCTGAGCCCGCTGGTACGTACGCCACC 9748  
QY 1103 ----- 1103  
Db 9749 GCAAGTAGTACACACAGATGCAATTTGTTCCCATGTCCCTTGCTACCTCCACCCC 9808  
QY 1103 ----- 1103  
Db 9809 ACCCAGAAACATGGGCTCTGCCCCGTACTGTAGGGTAGCAATTGCTTTTATATC 9868  
QY 1103 ----- 1103  
Db 9869 TAGTGGGCTCCCCCTGTATGTGTGTTTGTGTTTGGAGACAGAGTCTCACTCTG 9928  
QY 1103 ----- 1103  
Db 9929 TCACCCAGCTGAGTACATAGGCGAGTCTGGCTCACTGCAACCTCTGCTGAGT 9988  
QY 1103 ----- 1103  
Db 9989 TCAGCAATTTCTCTGCTCAGCCTCTGTAGTATCTGGAGCTCTGTAGTAGAGATT 10048  
QY 1103 ----- 1103  
Db 10049 ACAGGACCCACCAACGCCCACTAATTTTGTATTTTAGTAGAGAGGGGTTTCAAC 10108  
QY 1103 ----- 1103  
Db 10109 CATGTGTGAGTGTATCTGAACTCTGACCTGTGATCCACCGCTCAAGCTCCCA 10168  
QY 1103 ----- 1103  
Db 10169 AAGTTCTGGACTACAGGCGTAGCCACTGCGCCCCGCTGTGTATGTTTAAAGTCC 10228  
QY 1103 ----- 1103  
Db 10229 GTTTATCTTTGTGGAGACTGAAATGATGTGCATCCACATCTGAGTTTCTGAACACC 10288  
QY 1103 ----- 1103  
Db 10289 TACTATGTGAAGCATTTGTAATTAATAAATATGATATGCCAGTAGCTATGCTGT 10348  
QY 1103 ----- 1103  
Db 10349 AATCCAGTGTCTTAAGAGGCTGAGATGAGGATCACTTGAAGCCAGAAATTTGAGACCA 10408  
QY 1103 ----- 1103  
Db 10409 GCCTGACAAACAGCAAGACCCCATCTTACACACAAACAAAATTAGCCAGCATAGT 10468  
QY 1103 ----- 1103  
Db 10469 TGTGAGCACTGTGTCTCAGCTATTTGGAGTTGAGGTGGAGGATCGCTTGAAGCCA 10528  
QY 1103 ----- 1103  
Db 10529 GGAAGTTGAGGCTACAGTAGCTATGATCGTCCACTGCTCTCCAGTGTAGGCGACAGAG 10588

QY 1103 ----- 1103  
Db 10589 AGTCTGTCTTTAAAAAAGACTATCCCTTCTTGCCCTGTAAAGCATTT 10648  
QY 1103 ----- 1103  
Db 10649 CCTCTTCCAGGCCCGAGCTTTCATTTTGTAGCTAATAAATTCATTCCTATGTGA 10708  
QY 1103 ----- 1103  
Db 10709 TGTGATTCAGGTGTATGTATTTCTGGCTTAGGGTGTGTGGTAGATGAGACACC 10768  
QY 1103 ----- 1103  
Db 10769 AGCAAGTGTATTAGTGTAGCTTATAGTATAGAGTGTCTTCCGTGGGGCATAGCA 10828  
QY 1103 ----- 1103  
Db 10829 GGCATTTGGTGTATGTGTAGGCCAAGTGGTATGATAGTATAGAGCCCTAGAACCA 10888  
QY 1103 ----- 1103  
Db 10889 GAGAAAGACTGACTCAGTGGGCTCTGGGCTTCAAGTGAAGCTGCCAGCCCTT 10948  
QY 1103 ----- 1103  
Db 10949 CTATCTGTCCCTTCCGGTCCCCACCCGACTCACCTGTCTCCAGCATGGGGCAGTGGG 11008  
QY 1103 ----- 1103  
Db 11009 TGTATGTCACTGTATACGTGTACGTGCTCAGCTGATTTCTGTGGGTGGGGCCCCAGGCT 11068  
QY 1103 ----- 1103  
Db 11069 ACCCAGAGCCCAAGAGATTGGCGGAGAGAGATCTTTGTGTGATGTTCCATGGG 11128  
QY 1103 ----- 1103  
Db 11129 GATAGAGGTCTGCGGCCCAAGCAGCATGGGGCTCAGTGGGAGGGGAGCTAGGTGGCC 11188  
QY 1103 ----- 1103  
Db 11189 CACTGACCAATTACCACTCCCCCACCAGCACTCCCTGTGTGTGGGGAG 11248  
QY 1103 ----- 1103  
Db 11249 TGGGAAACAGCCATGTGTGCAATTCGTGTGTGTGATGTAGAGGCTGAGGGGCA 11308  
QY 1103 ----- 1103  
Db 11309 GGGGATGATTTTGTGGGTCTCAGGATGTCTGGGCTCTGTAGAAAAGTATGAAAAGA 11368  
QY 1103 ----- 1103  
Db 11369 CAACATAGGTTCCCGCGAGCAGACACACTTGGGGAAAGGAGCAGAGCGGTGC 11428  
QY 1103 ----- 1103  
Db 11429 TTGGGAGGAAAGGTGCAAAACATTAAGAGTGTGCCGTGCATCAGCAGACAAAGTT 11488  
QY 1103 ----- 1103  
Db 11489 TTGGAAAGATGCAAGAGACTCAATCTCATCCCTGAGCCCTGAAAGAGTATG 11548  
QY 1103 ----- 1103  
Db 11549 GCCCAGCTGACATACCTGGCTCAGCACCTTATGCCAGTTGAAAGAGAGCCTTTCT 11608  
QY 1103 ----- 1103  
Db 11609 TGTGTGTGTGCAAGAAATATGCAAGGCTGAATTTGACGCCCGAGGCTGGTCTGACTGTG 11668  
QY 1104 ----- 1119  
AspGlyValHisSerValIlnAlaIlnCysAlaLeuArgValIlnIle 1119

Db	11669	TCCCTCTCCACAGACCGGCTTACACACGCGACCGCCAGTGGCGGCTGGTGTCACATC	11728	Db	12749	TGGCATGAGAGTGTGGCTTGAATTAAAGGAAACGTAGATTGGGTGACTTGTCTCTG	12808
Qy	1120	IIethrAspGluMetLeuThrHisSerIIethrLeuArgLeuGluAspMetSerProGlu	1139	Qy	1319	-----	1319
Db	11729	ATCACCGATGAGATGTCTACCCACAGCATACGCTGCGCTGGAGGACATGTACCCGAG	11788	Db	12809	GTCTCATCACACCGGTTTAAGATGAGAGCTGGGCTTAGAACCTGTCTCTCCCCACTG	12868
Qy	1140	ArgPheLeuSerProLeuLeuGluLeuPheIIegIIaIaValIIaAIaThrLeuIIaThr	1159	Qy	1319	-----	1319
Db	11789	CCCTTCTCTGTCAACATCTGTAGGCTCTTCATCCAGGCGGTGGCCGCGACGCTGGCCAG	11848	Db	12869	AGCGAGTTCCTTTCTGAGCTGACAGAGTACTTTTCTCCCAAGGCAATTAAAGGGAGAGC	12928
Qy	1160	ProProAspHisValValIaIaPheAsnValIaIaArgAspThrAspIIaProGluGluHis	1179	Qy	1319	-----	1319
Db	11849	CAACCGACACAGGTGGTGGTCTTCAACGTACAGCGGACACGACGCCCCCGGGGCCAC	11908	Db	12929	ACAGAGAGCTGGGGGAACTGAGGGGTTGAGTTCTGAGATGGGTGCTCCAGAAAAC	12988
Qy	1180	IIeLeuAsnValSerLeuSerValIaIaGlnProProGluProGluIaIaProPhe	1199	Qy	1319	-----	1319
Db	11909	ATCTCTCAAGCTGAGGCTGTGCGGTGGCCAGCGGCCCGCGGGCGCGGCTTC	11968	Db	12989	ATTGACCTGATATGGGCTTGGGAGAAATGGCTCCGCTCCGCGCAGCTGTGTAAAGA	13048
Qy	1200	LeuProSerGluAspLeuGlnIaIaArgLeuIaIaLeuAsnArgSerLeuLeuThrIIaIle	1219	Qy	1319	-----	1319
Db	11969	CTGCCCTCTGAGAGACTGTGACGAGAGCGCTTATCTTCAACCGACGCTGTGACGGCCATC	12028	Db	13049	GGATCTGTGCTAGCGGCTTCTGTGTGGCTGTACAGCCAGGTCTTACATCTGTCT	13108
Qy	1220	SerAlaGlnArgValLeuProPheAspAspAsnIIeCysLeuArgGluProCysGluAsn	1239	Qy	1319	-----	1319
Db	12029	TGGGACAGCGCGGTGTGCTTCAACGACATCTGTCTGGAGGCTGTGACAGAAC	12088	Db	13109	CCCTGGATGTCACTGTCTGTCTTATGTCTTACCCCTCCCTTCAGGCTCTGTGT	13168
Qy	1240	TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIIeIIaSerSer	1259	Qy	1319	-----	1319
Db	12089	TACATGGCGTGTGCTGTGCGGTCTTCACTCTCCGCGCCCTTCAATCCCTCTCTCC	12148	Db	13169	CTTCTCAGGAAACCCCTGTCTCCCTTCCCTGTGGGCTTCTTCTTCAATGCCAGGC	13228
Qy	1260	SerValLeuPheArgProIIeHisProValIaIaGluLeuArgCysArgCysProProGlu	1279	Qy	1319	-----	1319
Db	12149	TCCGGCTCTTCCCGGCCATCCACCCCGTGGAGGGCTGGCGCTGGCCGCGCCGCGC	12208	Db	13229	CCTGTCTCTTGTCTTCCAGGGCCATGCCCACTGTGCTGTGCAAGAGCATAGAGGCC	13288
Qy	1280	PheThrGlyAspTyrCysLeuThrGluValAspLeuCysTyrSerArgProCysGluPro	1299	Qy	1319	-----	1319
Db	12209	TTTCAAGGTGATCTTGTGACAGAACAGATGTGACCTCTGTACTCGGGGCTGTGGCCCC	12268	Db	13289	CACCTCTTCCCGCTGCATCTTCTGTGCCAGAACTTGCCAGACTTGACTTCTGTGTGT	13348
Qy	1300	HisGlyArgCysArgSerArgGlnIaIaGlyTyrThrCysLeuCysArgAspIaIaTyrThr	1319	Qy	1319	-----	1319
Db	12269	CACGGGCGCTGCGGACGCGGAGGGCGGCTACACTGCTGTGTGTGTGTGTGTGTGTGT	12328	Db	13349	CTGTAGGGGTGGGGGAGAGGGAGAGGAGCGGCTTGGGGGAGGGAGAGAGTCTTGTCT	13408
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12329	GGTGAAGCAAGGAGGAGGACTCATGTGGCCAGCCCTGGAAGGCTGACTGTGTGTGTGTGT	12388	Db	13409	CAGAGATGTGTGGGGGGGTTTGGAGGGGGGGGTTTAGACAGGCGAACAAGTTTCC	13468
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12389	ACAAATCAGAGCAAAATGCTGGCGGCTGCTCATTTCTTCCGAGTGAAGTGTGTGTGTGT	12448	Db	13469	ATGAAAACAAAGACCACTCATAGGCAACAGAGCTGAAACTTATCTTGAATCTGGGA	13528
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12449	CTGAGAGGTCAATAATCCGGCTGTCTCCCTGTGACAGACCCCACTGAGGAAGCAAGTGTGA	12508	Db	13529	ATGGGAGGAGAGAAAGGTGTGATTTCCATCTGATGTGTGAGTTTTCAGCTCTCTGT	13568
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12509	GGAGTCTTCTCTGCTGT	12568	Db	13589	TGAATCTTGTCTGTCTGT	13648
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12569	CTCCAGAGCCAGTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	12628	Db	13649	AGGTCTTGT	13708
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12629	TGAACAGTAACTTAACCTTTCGCGCTCGGTATTTGAATGAGATTAATGACCGTAC	12688	Db	13709	TCCCAAGCTTCTCTCATATCTTTCATTAACCAAGCAAAATCTAATGTGAAGCTGTGT	13768
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	12689	TTAGGTCAATAGT	12748	Db	13769	CCATCCCAAGGAGGAGAACTGATCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT	13828
Qy	1319	-----	1319	Qy	1319	-----	1319
Db	13829	GGAGGAGAGAGAGGCTCCCAAGCCCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT	13888	Db	13829	GGAGGAGAGAGAGGCTCCCAAGCCCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT	13888

QY 1319 ----- 1319  
 Db 13889 ATTATCCCTCTGGGGGCCAGACACCCCAAGTGTGTTGGTAGTGTACATCGAAGGC 13948  
 QY 1319 ----- 1319  
 Db 13949 TCCCGGAGTCCGGGCTTACAAACAAGTGTCCAGCAGGCTGTGCTTATGAACCCAAA 14008  
 QY 1319 ----- 1319  
 Db 14009 TGGGTAAGGAAAGTGTGGGGACTCCGAGAACTTTCAGGGAGCAGTGTGGAGCTG 14068  
 QY 1319 ----- 1319  
 Db 14069 TGGCGAGACGGGTCAGTGACCAACAGGGTCCATGAAAGGCTGGGACATAGTTC 14128  
 QY 1319 ----- 1319  
 Db 14129 TCAGAGAGATGACACAGAGAGAAATGGGATGCTGGAGATTGGAGATTACG 14188  
 QY 1319 ----- 1319  
 Db 14189 AGGGTCAAGGAGGCCATGGAGACTATCACTGGGTTGTGGGGGTCTCAGTTCCCTCCT 14248  
 QY 1320 ----- 1325  
 Db 14249 GTCTCAGGTACTGTGTACTCACTGCTTTCTCTTTCAGGTGAGCACTGTAGAGTG 14308  
 QY 1326 SerAlaArgSerGlyValArgCysThrProGlyValCysLysAsnGlyValThrCysValAla 1345  
 Db 14309 AGTGGCTCGCTCAGAGCGCGTTGACCCCGGGTGTCTGCAAGAAATGGGGGCACTGTGTCAAC 14368  
 QY 1346 LeuLeuValGlyValPheLysCysAspCysProSerGlyAspPheGlyLysProThrCys 1365  
 Db 14369 CTGCTGTGGGGGTTCAAGTGCATTCGCCATCTGAGAACTTCAGAAAGCCCTACATGC 14428  
 QY 1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGln 1385  
 Db 14429 CAGGTACACCGGCGAGCTTCCCGGCCACTCTTCATCACTTTGGCGGCTGGCGAG 14488  
 QY 1386 ArgPheHisPheThrIleuAlaLeuSer----- 1394  
 Db 14489 CGTTTCACTTCACTCCCTGCCCCCTC-GTGAAGTGGCTGGCACTGGGGTGGGAGTGGG 14547  
 QY 1394 ----- 1394  
 Db 14548 CCTGGTGGGCACTGAGAGTCCCTCTGTCTGTGTGTTGAAGTAAGGTCAAGCAGTAA 14607  
 QY 1394 ----- 1394  
 Db 14608 AGCGTGAAGAGGCTCGCGCAGAGCAGAGTCTGTCTTGAAGTTGCTGTGGTTCTGCC 14667  
 QY 1394 ----- 1394  
 Db 14668 AGCAGATCTCCCTCCCTGCGCTGTAGGTGGCGGGGTCTGCGATCAGGTGGTAC 14727  
 QY 1395 ----- PheAlaThrLysGlyLysArgPheGly 1402  
 Db 14728 CCCATTCCTGCCCCCATCCCACTCTGTTCAGTTTCCCAAGAGAGCGGCAAGG 14787  
 QY 1403 LeuLeuLeuTyrAsnGlyValArgPheAsnGlyLysHisAspPheValAlaLeuGlyValIle 1422  
 Db 14788 TTGCTGTGTACATATGGGCGTTCAATGAGAGCATGACTTGTGGCCCTCGAGGTATC 14847  
 QY 1423 GlnGluGlnValGlnLeuThrPheSerAla----- 1432  
 Db 14848 CAGAGCAGGTCCAGCTCACTTCTCTGC-AGGTATCAAGTTCCTCCCATCTTGC 14906  
 QY 1432 ----- 1432  
 Db 14907 CATCTTCAAAAGGCCCAAGTCTTTCAGGCCCTGACCCCAAGCACAATCTATCAGCC 14966

QY 1432 ----- 1432  
 Db 14967 AAATGTGGCCCAAGCCCAAGCACTGGGCACCCAACCTTCCATCTTACCCCTGCA 15026  
 QY 1433 ----- GlnLysThrThrThrValSerProPhe 1442  
 Db 15027 GGCATTCAGCTCACTGTGCTCTTCCCAAGGGAGTCAACACGAGTGTCCCATTC 15086  
 QY 1443 ValProGlyValLysSerArgGlyValIleThrValGlnLeuLysTyrTyrAsnLys 1462  
 Db 15087 GTGCCGAGAGATCAAGTATGGCCATGTGCACTAGGTGAGCTGAATACTACATTAAG 15146  
 QY 1462 ----- 1462  
 Db 15147 GTGGGTGAGAGGCAAGAGGTTGGGGGTCTGTCTTGTCCCTCAGTGTCTTACCAGC 15206  
 QY 1462 ----- 1462  
 Db 15207 CCTGAGTGGCACTTCCCTCAGGCTTGGGTGGCTGTGTCAGGCAATTCTGCTGAGAGA 15266  
 QY 1462 ----- 1462  
 Db 15267 AGGAGCGGCTGGAAAGGTGCCACTTGTGGGCAATTCACCCCACTGGGCACTGCTG 15326  
 QY 1462 ----- 1462  
 Db 15327 ATGCTCTGTTCACACTAGGCAAGCGGAAACTTGGACACTGAGGAAGTACTATATGCA 15386  
 QY 1462 ----- 1462  
 Db 15387 TAAACACGTAGGACACGTTGTGTATATGCAATGTGTACGTTGTCTGGGCAATGTG 15446  
 QY 1462 ----- 1462  
 Db 15447 GTGCACACAGGCAACCAAGTGTGCTGTGCGCCATGAACCTATGGGCTCTGTACTCAGTG 15506  
 QY 1463 ----- ProLeuLeuGlyGlnThrGlyLeuProGln 1472  
 Db 15507 TCCCCCTTCTGTGCTCTTCTCTGTGTCACAGCACTGTGGGTGAGACAGGCTCCACAG 15566  
 QY 1473 GlyProSerGlyLysValLysValAlaValThrValAspGlyCysAspThrGlyValAla 1492  
 Db 15567 GGGCATTCAGAGCAGAAAGTGGCTGTGTGACCGTGAATGCTGTGACACAGAGTGGCC 15626  
 QY 1493 LeuArgPheGlySerValLeuGlyAsnTyrSerValAlaGlnGlyThrGlnGlyVal 1512  
 Db 15627 TTGGCTTCGGAATCTGTCTGTGGCACTACTCTGTCTGTGCCAGGGCAACCAAGGTGGC 15686  
 QY 1513 SerLysLys----- 1515  
 Db 15687 AGCAAGAAATGAGCAGGGGAAAGGCCAGAGATGGGTGAAGTGAAGGCAAGGCCAGGCA 15746  
 QY 1515 ----- 1515  
 Db 15747 GGAAGGAGGTGAGAGGCTGTTGGGTTCTCTCTCTTGTGCTCTTAAAGCC 15806  
 QY 1515 ----- 1515  
 Db 15807 CTGCTCAGCCCTGGGGATGGGGTCAAGACTGGGGAATCAGAGCAAGGCCCTGTGGGAG 15866  
 QY 1515 ----- 1515  
 Db 15867 TCTGGAGAGGGGCTGAGTGTTCAGAAAGAACATTCAGAGTAAAGGATGAGCCCTTCT 15926  
 QY 1515 ----- 1515  
 Db 15927 GAGCACTGCTGTGAGAGGCTCTTGGACACTGTGACCGCTTGTGTCTGTGGCTGG 15986  
 QY 1515 ----- 1515  
 Db 15987 CGGACAGAGCGAGAGGACCAAGAGCTCCAGGCCCTCTTGTGCGGCACATCTTGG 16046  
 QY 1516 ----- SerLeuAspLeuThrGlyProLeuLeuLeuGlyValProAspLeuPr 1532

Db	16047	CTCCCCCAGGCTCTGATCTAAGCGGCGCCCTGCTACTAGCGGGGCTCGACCTGCGC	16106	Db	17127	GAGGACATATAGAGCGCGCTGATCCGTTGGGAAAGTCAATGCTGCCACTGTGGGCTT	17186
Qy	1532	oGlusPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAs	1552	Qy	1666	-----	1666
Db	16107	CGAAGGCTTCCAGTCCGAGATCCGAGATTCGTGGGCTGCAAGCGAACTCTCAAGTGA	16166	Db	17187	GAGGAAATTAATACGCTTTCCTCTATAGGCTTCAGCTGGGCAAGGGATGGTTGAT	17246
Qy	1552	pSerArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrVal-----	1568	Qy	1666	-----	1666
Db	16167	CAGCCGGCACATAGCATGGCTGACTTCATTCGCAACATGACACCGTGCCTGTATGGG	16226	Db	17247	GATCCAGAAAGGGAAGTGTAGCTGGAGAAACAGTATTAAGCCGCTGTGACCTGT	17306
Qy	1568	-----	1568	Qy	1667	-----GlnLeuArgGlyGlyIleIleValMetLeuSerValGlnGlyThr	1680
Db	16227	GGCCCGGGGTGAGCCAGGCTGGATCCAGTGTAGAGAGAGGGGCTGTGCTGTGTGG	16286	Db	17307	TTTCTTTCCTCTGT	17366
Qy	1568	-----	1568	Qy	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyValArgAlaAsnAspGlyAspTrpHis	1700
Db	16287	AGGATGCCAGATGAGGGGTCTCAGAGGCTCTCTAATCTGTGCCACAGGCTTCTCTTT	16346	Db	17367	GGGCTTCAAGGCT	17426
Qy	1569	-----ProGlyCysProAlaIleVal	1575	Qy	1701	HisAlaGlnLeuAlaLeuGlyAlaSerGlyValProGlyIleAlaIleLeuSerPheAsp	1720
Db	16347	AGAGCCCAAGCTCAGGTCCAGTCAAGCTGCTCTCTCCATTCAGAGCTGCTGCGCAAGA	16406	Db	17427	CATGCAAGCTGGCACTGGGAGCAGCGGGGGGCCCGGCGCATGTCATCTGTCTTCAT	17486
Qy	1575	-----	1595	Qy	1721	TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer	1740
Db	1575	YSAsnValCysAspSerAsnThrCysHisAlaAsnGlyValThrCysValAlaAsnGlnTrpAspA	1595	Db	17487	TATGGGCAAGCAAGAGCAAGAGGCACTGGGCGCCCGGCTGATGCTGTGCACTTGAAGC	17546
Qy	16407	AGAACGTGTGTACAGCAACACTTGTCCACATGGGGGCACTTGTGTGAACAGTGGAGCG	16466	Qy	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly	1760
Qy	1595	laPheSerCysGlnCysProLeuGlyPheGlyGlyValSerCysAlaGln-----	1611	Db	17547	AACATAACAGTGGCGGAAATACCTGGGCAAGCGGCGGTGTGGCTTGGCTTGGGGC	17606
Db	16467	CGTTAGGCTGCGAGTGGCCCTTGGGCGGCAAGCTGCCGCCAGGGGTAGAGAGG	16536	Qy	1761	Cys-----	1761
Qy	1611	-----	1611	Db	17607	TGTTTGAGGTAAGTGTCTGCGCTGCTCCATCCCTCCCGCCACACCTGAGCGCC	17666
Db	16527	GCGGCTGTAGAGGCCACAGCTGCGGTGCATCAACATGTGGAATGCGAGGGTTGGG	16586	Qy	1762	-----Leu-GlnGlyValArgValSerAspThrPr	1771
Qy	1611	-----	1611	Db	17667	GCTCCACTGAGGGCAACATGCTCTGTCTCTCCCAAGGAGTGTGGGAGTGAAGCAAGGCC	17726
Db	16587	GCAAGCACTGGGCAAGGCTCTGTGAGCGGGCTGTGGGTGAAAAAGCTGTCTGGGCGAG	16646	Qy	1771	oGlucIleValAsnSerLeuAspProSerHisGlyGlnSerIleAsnValGlnGlnGlyC	1791
Qy	1611	-----	1611	Db	17727	GAGGGGGTTTAACAGCTGTATCCACCATGGGGAAGATCAACGTGAGAGCAAGGCTG	17786
Db	16647	AGCCCTGAAGAGGCAACAGAGAGCAAAAGGGCCAGTCAAGAGCGCTGGGCAAGC	16706	Qy	1791	sSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerIleCysSerAsnAs	1811
Qy	1611	-----	1611	Db	17787	TAGCTGCTGACCCCTGTGTATCTCAACCGGTGCTGTATCAACAGTATTGCAACAAG	17846
Db	16707	GTTGGGCGGTCCTCCAGGCGAGTACGCACTTTGAGAGCGGGGCTGATGAGGGAGTGGGC	16766	Qy	1811	pThrAspSerIleThrSerCysAsp-----	1820
Qy	1611	-----	1611	Db	17847	CTGGAAAGCTAATCTGTGAGCTGTGA-TCAGGTATGTAAAGATCAGGGCAAGGGC	17905
Db	16767	TCTGCTCTCCGTGAGCGCGCAACGCTGAGCATCAAGCCCCAGGGCCCTTGAAGCTCC	16826	Qy	1820	-----	1820
Qy	1612	-----GluMetAlaAsnProGlnHisPheLeu	16826	Db	17906	AGTTATCAGGTGCTGGGGCCACATGCTGGCTGTGTGTGTGTGTGTGTGTGTGTGT	17965
Db	16827	ACCCTCACAAGTCTGCTTTCCTGCTCCCGCAAAATGGCCAAATCCACAGCATCTTCTG	16886	Qy	1820	-----	1820
Qy	1621	GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTrpLeu	1640	Db	17966	TCAACACCTCTCTGAGTCTTAGTTGCTTACACAACAAAGCATAGCAGCTGATAGGA	18025
Db	16887	GCGACACACCTGTGGCTGTGATGGCTCTGTGCTGCCATCTCCCAACCTGTGTACTTC	16946	Qy	1820	-----	1820
Qy	1641	SerLeuMetPheAsnArgThrArgGlnAlaAspGlyValIleLeuGlnAlaIleThrArgGly	1660	Db	18026	TGCTATGTGAAAGAAACAGATGAAGGTTTGAATGAAAGTGTGCCAGGGAGGGTGGG	18085
Db	16947	AGCTCTCATGTTCCGACCGCGCAAGCGCAGCGGTGTCTGTGCGAGCATCAACAGGGGG	17006	Qy	1821	-----ProGlyIleValArgValAspAsnGlyThrAsnVal	1831
Qy	1661	ArgSerThrIleThrLeu-----	1666	Db	18086	GGAATGAGCCTCTGTGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	18145
Db	17007	CGACACACATCAACCTCTACAGGTATGCAATGGAAGGGGCGCTGGCCCTGCGCTGCAATA	17066	Qy	1831	lCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgIleProSerAlaProHis	1851
Qy	1666	-----	1666	Db	18146	GTTGACCTGAACCGGTGTAGCAACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	18205
Db	17067	GAGCCCTGTGAGCTCTAGAGCGGCTGACAGAAATGCTGGCAGAGTCTGGGCGAGGGGT	17126	Qy	1851	sGlyIleThrCysGlnCysProProAsnIleLeuGlyProTrpCysGlnIleThr	1868
Qy	1666	-----	1666	Db	18206	TGGCTAATCCTGGAAGTGTCCCGCAATTAACCTTGGGCGCATCTGTGAGACAGGTAAAC	18265

QY 1868 ----- 1868  
 Db 18266 AGACCAAGGCGATGTGGCAGCAGAGTCCAGTGGCTGTCTTCTCTGTGTGTCTCCCTCA 18325  
 QY 1868 ----- 1868  
 Db 18326 GAGCCCCGAAAGCCTGGCTGATTCACAGCCAGGGTCAAGAGGCCACATAGGGCTCACT 18385  
 QY 1869 ----- Arg11ea 1871  
 Db 18386 AGCTTAGTGGAGATGGCAGGGGAGCTCATGCTTACCTGGCTCTCTGTGCAAGATTG 18445  
 QY 1871 SPGLNPROCYSPROARGLYTTPTRGLYHISPROTHRCSGLYASAPY 1891  
 Db 18446 ACCAGCTTGTCCCGGTGGCTGTGGGAGCATCCCAATGTGGCCCACTGATG 18505  
 QY 1891 A1SERLYSGLYPHAPPROAPCYASAPLYTHSERGLYGLUCYSHISGLYVS--- 1909  
 Db 18506 TCAGCAAGGCTTTGACCCAGACTGCACAGCAAGCGCGAGTGCCTGCAAGTGA 18565  
 QY 1909 ----- 1909  
 Db 18566 CAGCCCCAAGCAAGCTCCACTGTGGCCACTTGGGCTCTGTCCACATCTCTGGGCTT 18625  
 QY 1909 ----- 1909  
 Db 18626 AGAGGGCAGGCTGTAGAGATGGGTCTTCTGTAGAGCTGAGGGCTTGTGCTTGTCTCT 18685  
 QY 1909 ----- 1909  
 Db 18686 CACAGGAGGCGAGTGTGCTGTACCTTCTGTGTCTGAGGCGCAGGGGACCCCTGGGC 18745  
 QY 1909 ----- 1909  
 Db 18746 CAATGTGTAAACCATAGCTCTGGCTGTGGTGTGGTGTGGTGTGGACCTGTCTGATTC 18805  
 QY 1909 ----- 1909  
 Db 18806 TGTGACATCTCGCTGATGGGCTCACTCCAAAGCATTTGATGTAATTATTATTGATA 18865  
 QY 1909 ----- 1909  
 Db 18866 CTGCTCTGAGGGGACCTGAGTGGGACCTCTCCACAGGGGCCACAGCTGAGAAAGAC 18925  
 QY 1910 ----- G 1910  
 Db 18926 CCATGACATCTCGGAGAGAGAGCCTGGCCAGTGAACCGTTCTTCTCTTCTCAAG 18985  
 QY 1910 LUBANH1STYRARGPROGLYSEPTHTCYALEULEUCYAAPCYSTYRPROTHR 1930  
 Db 18986 AGAACCACTACCGGCCCCCAGAGGACCCCACTGCTCTTGTGTGACTGTACCCCAAG 19045  
 QY 1930 L1SERLEUSEARVALCYAAPPROGLUBANGLYGLINCYSPROCYSLYSPROGLYVALI 1950  
 Db 19046 GCTCCTTGTCCAGAGCTGTGACCTGAGGATGGCCAGTGTCCATCCAAAGCCAGGGTGA 19105  
 QY 1950 L1EGLYARGLYNCSAAPARGCYAAPASAPPROPHELAGLYVALITHTHRAAGLYCYEG 1970  
 Db 19106 TGGGGGTGATGTGACCGCTGTGACACCCCTTTCTGAGGTCAACCAATAGGTGTG 19165  
 QY 1970 L1----- 1970  
 Db 19166 AAGGTGGGCTCTGTGGATGGGTGGGCAAGCCCTCTTAAGTGTGTAGGCACTGTACAC 19225  
 QY 1970 ----- 1970  
 Db 19226 CCAAGAAAGCCAGAAAGGGCTGTGTGCAAGCCTGGGGAGGCTCTGGGAGGCTTAGGGAG 19285  
 QY 1970 ----- 1970  
 Db 19286 GGCTGGGAGCTGTGGCAGGACGACAGAGGAGAGCTGTCTCTGTGGTGAACATGTCTCT 19345

QY 1971 ----- ValAsnTyzAspSerCysProArgAlaIleGluAlaGly1IleTPTRPro 1988  
 Db 19346 TCCCGCAGTGAATTAATACAGCTGTGCCAGAGGATTTGAGGCTGGATCTGGTGGCCCC 19405  
 QY 1988 THTHARGPHEGLYLEUPROAlaAlaAPROCYSPROLYSGLYSErPhe----- 2004  
 Db 19406 GTACCCGCTGGGGCTGTGCTCTGTCTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 19465  
 QY 2004 ----- 2004  
 Db 19466 GAGGCCCCGATGTGATGTGAGAGACATGCGCTCTGTCTGTGATGGATGAGGCGAGAA 19525  
 QY 2004 ----- 2004  
 Db 19526 GCTCCTGGCTGAAGATCCGGGAGCCCACTTCTGGCTCCAGGTCCTGGATTCAATT 19585  
 QY 2004 ----- 2004  
 Db 19586 CCCATCTTCTTGGTTCATATCCCTTCCCATGTGGCCAGGTGTGCTTACATCTTG 19645  
 QY 2004 ----- 2004  
 Db 19646 GACTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 19705  
 QY 2004 ----- 2004  
 Db 19706 CAGGCTGTGTGAGTGTGGCGTATCTCCGCTCGGTGCAACCTGTGCTCCAGGTTCAAG 19765  
 QY 2004 ----- 2004  
 Db 19766 CAATTTCTGTCTTAAGGCTGTGCGCCGACACACACTGTGCTAAATTTTTTTTTTTTT 19825  
 QY 2004 ----- 2004  
 Db 19826 TTTTTTTTGTATTTTGTATGAGACAGGGTTTCAACATGTTGGCCAGTCTGTCTCGAA 19885  
 QY 2004 ----- 2004  
 Db 19886 CTCTGACCTGATATCCGCTGTGCGCTCCCAAGTGTGGATTAAGGCGGTAG 19945  
 QY 2004 ----- 2004  
 Db 19946 CCACTGACCCCAACCAATTTTGAATTTTTTTTTTTTTTTTGTAGACGAGTCTGCTC 20005  
 QY 2004 ----- 2004  
 Db 20006 TGTATCAGGCTAAGATGAGTGTGATGATCTCGGCTCTGCAACCTCACTCAAG 20065  
 QY 2004 ----- 2004  
 Db 20066 TTCAAGCATTTCTCTTACCTGAGCTTCCAGTAGCTGTACTACAGGCGTGTCTACA 20125  
 QY 2004 ----- 2004  
 Db 20126 TGCCAGCTAATTTTTTATTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 20185  
 QY 2004 ----- 2004  
 Db 20186 TGTGTGATCTTTGACCTTGATGATGACCGCGCTGTGCTCCAAAGTGCAGATT 20245  
 QY 2004 ----- 2004  
 Db 20246 CAGGCTGAGCAACGTCGCCAGCCACATCTTGACTTTAACAATAATCTGAATAAC 20305  
 QY 2004 ----- 2004  
 Db 20306 CTGCACTGTGCTTGGCATAGGGCTGTGTAGAGATTCAAGATGATCATGAGGTTCC 20365  
 QY 2004 ----- 2004  
 Db 20366 AATGTGACTACCAACCAAGTCTGTGAGTTACGCCACCAAAAGCAGACAGTCTGGGG 20425  
 QY 2004 ----- 2004

```

Db 20426 AGTAGGGGGTGTGGCTGCTGAGCCAGAGACACCCAGCCAGAGATGGGGGCTGGCGCTGGT 20485
QY 2004 ----- 2004
Db 20486 CCTGAGCTGGCAATGAGCAGGTTTCATAGGAGAGAAATGGCGGAGAGACAAAGTGTGGC 20545
QY 2004 ----- 2004
Db 20546 TGGAGACAGGCAGAGATGGGGGCTCAAGAGAGTCGTATGCCCTGTTCCCTTAGAGCCCATG 20605
QY 2004 ----- 2004
Db 20606 TGTGAGCTGATCTTGGGAGAAATTTTGGCCCTGGTGTCTACATACCTGTCTTCTGTTT 20665
QY 2004 ----- 2004
Db 20666 CACAGGGGCGCTTGTGTTGTAGTTACTGTCCCGGCCCATGAGGTGGCTCCAGCAG 20725
QY 2004 ----- 2004
Db 20726 AACCTCTCAGCCAGTCTCAGGGGCTTCTGTGTCTCTCCCTGAGGCGGCGCTTTTGGCT 20785
QY 2005 ----- 2005
Db 20786 TCTTCCCCCAGAGGACTGTGTGGCGCACTGTGATGAGCAGAGGGGGTGGCTCCCCCAA 20845
QY 2021 snlLeuPheasnCythrSerIlethrPheSerGluLeuIysGlyPhe----- 2036
Db 20846 ACCTTTCACTGACAGCTGCATCACCTTCTCAGAACTGAGAGGCTTCTGTAAGTGAACCC 20905
QY 2036 ----- 2036
Db 20906 CTGATCTCCATCTTTTCCCTGTCTTCCTGACGTGCTACCTTGGCCCCCATCCCATCTT 20965
QY 2036 ----- 2036
Db 20966 TGAAGACGGGGCTTCTGGAATTCAGCCTGTGTGTCTGGGCTCCAGCTGGAGAGGCC 21025
QY 2036 ----- 2036
Db 21026 GTCTCTCACTCTGAGGTCCTTTTGTCTCAGAGTTCCTGCCCTCAGCTCTGCTCTCCG 21085
QY 2037 ----- 2037
Db 21086 CTCCGTCCTGTCTAGGCTGAGCGGCTACAGCGGAATGAGTCAAGGCTTCAAGGCGGC 21145
QY 2052 SerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnIleThrAlaGlyThrPheGly 2071
Db 21146 TCCGAGCAGCTAGCCCTGCTCTCGGCAACGCCAGCAGCAGCAGCAGCTGACTTTCGGC 21205
QY 2072 SerAspValIysValAlaIleArgIleLeuAlaThrArgLeuLeuAlaIleGlnSerThrGln 2091
Db 21206 AGCGAGGTCAAGGTGGCTTACCAAGCTGGCAGCGCGGCTGTGGCCCAAGAGAGCCAG 21265
QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValIlePheThrGln----- 2106
Db 21266 CGGGGCTTTGGGCTGTCTGCCACACAGGACGTGACTTCACTGAGGTGGGGCTTGGAGGA 21325
QY 2106 ----- 2106
Db 21326 TGCAGGGCTGGCTGTAGATAGAGGGGTCAATGGTGAAGAGCTGCTCATATGGCACTGGG 21385
QY 2106 ----- 2106
Db 21386 GGGCAGAGGGGGCGCTCCCATCCACCTAACAAGAGGCTCCCTGTGGAGAGAGGCTTCCA 21445
QY 2106 ----- 2106
Db 21446 TGAACCTGTGACCCCTGGGCCAGACCTCTTCCCAACCTTCTCATATGAATATATGTG 21505
QY 2106 ----- 2106

```

---

```

Db 21506 TGTGGGCAATGTTCTGACCCAGAGCAGAGGCTGCTGTGGTACCATAGGCTGGAGGA 21565
QY 2106 ----- 2106
Db 21566 AGCAGTATCTCAGACATGTGTGTGGTTCAGGACAGAGGCTGGAGACTTATATGTAGA 21625
QY 2106 ----- 2106
Db 21626 GAATGAGAGAGGGCTTAGGGGGCAGAGCATTCAGTGACTGAGTGGGCAATGGGCTTCA 21685
QY 2106 ----- 2106
Db 21686 GAGCATTTCAAGGGAGCAACAGTGAACACAGAGGAGGGGCTGTGGGAGCAGGGA 21745
QY 2106 ----- 2106
Db 21746 CTGGCCGGGACAGGCGGGGCTTCATGTGGGATCTCAGGTGTGGGTTGTGACACA 21805
QY 2107 ----- 2107
Db 21806 CCCACCGTAGCCTTGCCCAACCCAGAAATCTGCTGGGGTGGCAGCGCCTCTGGACAC 21865
QY 2118 rAlaAsnIysArgHisTyrGlyLeuIleGlnGlnThrGlyLeuIleGlyThrAlaIlePheLeu 2138
Db 21866 AGCCAAAGGGGCACTGGGAGCTGATTCAGCAGACAGAGGGTGGCACCGCCTGGCTGCT 21925
QY 2138 uGlnHisTyrGlyAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSe 2158
Db 21926 CCAAGCATATAGAGGCTTACGGCAAGTCCCTGGCCCAAGAACATGGGGCACACTTACTTAG 21985
QY 2158 rProPheThrIleValThrProAsnIle----- 2167
Db 21986 CCCCTTCAACATTCGTCAAGGCCCAACATTGTAAAGCTGTGCTGGGTGGGAGGGGTT 22045
QY 2167 ----- 2167
Db 22046 TGTGAGGAGTCCCGACAGAGCGGCTGTGCTGGGTCTGCTGCTCAGAGTCCG 22105
QY 2168 ----- 2168
Db 22106 ATCTGTGACATCCCTTCTTACTCATCTCCGATGGTGGCTTGGACAAAGGAATTGG 22165
QY 2180 lAgIyAlaIysLeuProArgTyrGlyAlaLeuArgIlyGlyGlnProProAsnLeuGlyT 2200
Db 22166 CTGGGGCCAGCTGCCCGCTACGAGGCGCTGCGTGGGGAGCAGCCCGGACCTTGAGA 22225
QY 2200 hTrhValIleLeuProGlnSerValPheArg----- 2210
Db 22226 CAACAGTCAATTCCTGAGTCTGTCTTCAG-AGTCAATGGTGGCCATGGATTGAAGTTG 22284
QY 2210 ----- 2210
Db 22285 GGAAGTGAACCCAGTGTCTGTGACACTCCACAGAGAGCAGGGGCCAGCTAATGTGAC 22344
QY 2211 ----- 2211
Db 22345 AGTGTCCCTTCCAGAGACCGCCCGTGTGCAAGGCCGCGAGGCCCGCGAGAGGCCAGG 22404
QY 2226 lProGlnGlnLeuAlaArgArgIleArgArgHisPProGlnLeuSerGlnIlyGlnIly 2246
Db 22405 AGCCAGAGAGGCTGACCGGGAGACAGCGCACCCGAGCTGAGCCAGGGGTAGAGGTG 22464
QY 2246 aAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProA 2266
Db 22465 TGGCCAGGTCAATCATCTTACCGACACCTTGGCGGGCTTACTGCCCATATATACCTTG 22524
QY 2266 sPlyAsrSerLeu----- 2270
Db 22525 ACAAGCCAGCTTGAAGTCAAGCTTAGGGGACAGGTGTGGTAAAGGGTATAGGTCGGGC 22584
QY 2270 ----- 2270
Db 22585 GGTAGTGTGAGGCAATGAGAGGGGCTTCTCCCAATGATGACTGCCGTGG 22644

```

QY	2271	-----ArgValProIaArgProIleIleAanThrPro	2281
Db	22845	TGACTGTGCACCTGACTGCCCCACATCAAGATGCCCAAGCGCCGATCATCAACACAGCC	22704
QY	2282	ValValSerIleSerValHisAapAapGluGluIleuLeuProArgAlaLeuAapLysPro	2301
Db	22705	GTGGTGAAGCATCAACGCTCCATGATGATGAAGAGAGCTTCGCCCGGGCCCTCGAACAAACC	22766
QY	2302	ValThrValGlnPheArgLeuLeuGluThrGluIuArgThrLysProIleCysValPhe	2321
Db	22765	GTCAACGGGTGACGTTCCCGCTGCTGGAGACAGAGGAGCGAACAGCCCATCTGTGCTTC	22824
QY	2322	ThraAnHisSerIleLeu-----	2327
Db	22825	TGGAACCATCAATCATCTGTGAGAGCTGCACATGCCCTGCGCCCTCAAGGCTTCGGGCTGA	22884
QY	2327	-----	2327
Db	22885	GTCCAGGCCCTCGCATGCTCAACCTCCTTGTCCTCCCTGACCTGCTTCCTCCACAG	22944
QY	2328	ValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAenGlySer	2347
Db	22945	GTCAGTGGCAACGGGTGCTGGTCGGCCAGAGGCTGTGAATCGCTTCCTCCGATGAGAC	23004
QY	2348	HisValSerCysGlnCysAanHisMetThrSerPheAlaIleuMetAapValSerArg	2367
Db	23005	CACCTCAGCTGCCAGTGAACCATGACATGACAGAGCTTGCTGTGCTCATGAGACGTTTCTCG	23064
QY	2368	ArgGlu-----	2369
Db	23065	CGGAGGATCGGGCCCAAGGGGCACTGACAGAGCCGCGGGTGGGCAACCAAGGCAAGGGG	23124
QY	2370	-----As	2370
Db	23125	CTGGGTGCTCAGGTCTGACCTTCCTAATTCCTCGAGCCCTCGCACTACTCGACAA	23184
QY	2370	ngIvGluIleLeuProLeuLysThrIleThrTyrValAlaLeuGlyValAlaThrIleuAlaI	2390
Db	23185	TGGGGAATCTGTCCATGAAAGACATGACATACAGTGGCTGAGGTGTCACCTGTGCTGC	23244
QY	2390	AlaLeuLeuLeuThrPhePhePheLeuThrIleLeuLeuArgIleLeuAArgSerAanGlnHisGln	2410
Db	23245	CTTCTGCTCATCTTCTTCTTCTCCATCTCTCTTGCGATCTCGCGCTCCCAACCAACAGG	23304
QY	2410	YIlaArgArgAenLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyI1	2430
Db	23305	CATCCGACGTAACCTGACAGCTGCCCTGGGCTGGGCTGACACTGATCTTCTCTCGGGAAAT	23364
QY	2430	eAnGlnAlaAapLeuPro-----	2436
Db	23365	CAACCAAGGCTAACCCTCTGTAAAGATCTCTACTGCCCAAGAAACTGTCCCCACCTTCTC	23422
QY	2436	-----	2436
Db	23425	AGGCGGCTCCCGACGCCCACTGGAACCCGTCCTCTGACCATGAATCTTAATAAG	23484
QY	2436	-----	2436
Db	23485	TGCCTAGTGACACACTGGGCCAGAGGTTCTCTTCTGTGGCTCCCCCGGATCCCCAG	23544
QY	2436	-----	2436
Db	23545	CACCTGCTCTGGGCCCAAGGCTTCCTGGAAGACAGTTCACCAACACCAAGGCCCTCTCAT	23604
QY	2436	-----	2436
Db	23605	GCTGAGCCCGAAGCAGAGCTGTGCTCTGGGCGGGCCCGGTCGTGACCTGCGCTGCC	23664
QY	2437	-----PheAlaCysThrValAlaIleAlaIleLeuLeuHisPheLeuTyrLeuCysT	2453
Db	23665	TGGGCCCTCAGTTTGCTCGACAGTCAATGCAATCCGTGTCGACCTTCTGTACTCTCTGCA	23724

OY	2453	hrPheSerTPRALaleuLeuGIuAlaleuHIsleuTYRAspAlaleuThGIuValALysA	2473
Db	23725	CTTTTCTGGGCTCTCTGTGAAGGCTTGACCTTGACCGGGACATCACTGAAGGTGGCG	2378
OY	2473	spValaenThngIyPrometArgPheTYTThleLeuGIyTPRGIyValProAlahei	2493
Db	23785	ATGTCACACCGGGCCCATGGCTTCTACTACATGTGGGCTGGGGGGTGCCTTCATCA	2384
OY	2493	leThr-----	2494
Db	23845	TCACAGGTACTCCACCCCAATTCACAGTCTTGGGGTCCACATCCCTGGGTTCACCTTGT	2390
OY	2494	-----	2494
Db	23905	GCATATGTTCTTCACCCCAATACAGGACCTTGAGGGCCCAATCCCATGGCCAGGCG	2396
OY	2494	-----	2494
Db	23965	CCTTATTCACAGGTGTCCCTGTGGTTTAAACCAACTGTGACGGCCACCAAGGCGTCA	2402
OY	2494	-----	2494
Db	24025	CTCCCTTATCTGGGGGAGATGGGAGGGGCGCATGTGGGAGAAACCTTTTCATGCTT	2408
OY	2494	-----	2494
Db	24085	CATGCTGGCCCTGTGAGCCCACTGGCCGAGCCCTACTTCCAGGCCCTCATCACCC	2414
OY	2495	-----	2504
Db	24145	CTCCACTGCTCCCGCTGTGTCTCAATGCTCAGAGCTTA-GCCGTGGGGCTGAACCCCAAG	2420
OY	2504	lyTYRGIyAenProAspPheCyetrPheSerIleTYRAspThIleuIleTPSerPhea	2524
Db	24204	GCTACGGGAACCTGCACTTCTGCTGGGCTCTCCATCTATGACACGCTCATCTGAGATTGG	2426
OY	2524	IaGIyProValAlaPheAlaValSer-----	2532
Db	24264	CTGGCCGGGTGGGCTTGTGGCGTCTCGGTGAGTGAAGTGAAGTGGTGGTGTGACCTGT	2432
OY	2532	-----	2532
Db	24324	GGCCCTCTTGTGCTGTCTCTGTGCTGTGGGGCCGGGTGTCTCAGGACTCCTTTGA	2438
OY	2532	-----	2532
Db	24384	ACAAGTGAAGGCAACGGGGGCTGTGGGGCTGGGTGAAGCTGTGTGCCCATGAGACC	2444
OY	2533	-----	2545
Db	24444	CSATGCCCTGTGCCCTGTGCTAAGTGAATGTCTTCTCTGTACATCTGGGGGCCCGGGCT	2450
OY	2545	erCyValaIaIaGIaTArgGIuGIyPheGIuTyIyAluGIy-----	2557
Db	24504	CTGTGTGTGGCAGCGGCAAGGCTTTGAGAAAGTAAGTCTGTGTGATGAAGGTGGG	2456
OY	2557	-----	2557
Db	24564	GTGCTGGGCAATGGGAGGCAACCAAGTGAAGTCTATGACCGGCACTATGACCTGT	2462
OY	2558	-----	2575
Db	24624	CCCTATGCCCAAGC-TGGGGCTTGAGGCTCTTCGGCGGCTCTCTGTCTGTACAGGCCA	2468
OY	2575	hTTPLeuLeuAlaLeuLeuSerValaenSerAphThIleuLeuPheHieTYTLeuPhea	2595
Db	24683	CGTGGCTGTGGCACTGCTCTGTCAACAGGCAACCCCTCTTCCACTACTCTTTGG	2474
OY	2595	IaThIyAenCyAlle-----	2600
Db	24743	CTACTCTCAATGTGCATCAAGGTACTGTGGCCAGCCTGTGAGAAAGGAGGCACTGGGCT	2480
OY	2601	-----	2606



Db	24803	GTGATGCGCTGAAATATGACACAGACCGTGTGCTCTTCTTGCGTGCAGAGGCCCTTATCT	24865
Oy	2606	helauSerTyrValValLeuSerIysGluValArgLysAlaLeuIleuAlaCysSerA	2626
Db	24863	TCCTCTCTTAAGTGTGTGTAGCAGAGAGGTCGGAAAGACATCAAGCTTGGCTTCGACGC	24922
Oy	2626	IglYsProSerProAspProAlaLeuThrThyIysSerThrIleuThSer-----	2642
Db	24923	GCAAGCCAGCGCTGACCTGTCTTGACACAAAGTCAACCTGACCTCGGTAGGAGAC	24982
Oy	2642	-----	2642
Db	24983	CAGGGGTCTCAGAGGCGGTGAAGAGAGGAGAGGAGGCGCCAGTGAAGCTGTACT	25042
Oy	2642	-----	2642
Db	25043	TTTGGCCCCATCCCTCTTCTCTTTTCCATCCCTTCTTGAAAGTGAAGGAGAGGTG	25102
Oy	2642	-----	2642
Db	25103	AAATGCTGTTCGGCCTGGGAGAGATGTGGAAAGGTGTGTAGATTGTAATTGTA	25166
Oy	2642	-----	2642
Db	25163	GGGAGATGGGCTGAGACAGATTAATCAGTAAATAGTCCTTGCACTGAGAACTCTCA	25222
Oy	2642	-----	2642
Db	25223	TGAATTTTATGTAATTTTCTGTCTCCCAAGTGTCTCAGAGACAAATTTCTGTCTCAT	25282
Oy	2642	-----	2642
Db	25283	CTTTTACAGACCATTAATATAGTAGTCAGAGATCCCTGGGGCATGACAGGTTAGACA	25342
Oy	2642	-----	2642
Db	25343	GCTGTGGCAGAGCTGTACCACTGCCAGGCTCAATATGGGGACAGCGAAGACAGGGC	25402
Oy	2642	-----	2642
Db	25403	CTTCTCTTTGACACTGAGCATGTGAAGGTGGAATTAGGGGCAAGTCCCTCACCTTC	25462
Oy	2642	-----	2642
Db	25463	CTTCTCCAGCGCTGGGGCCAGCCATCCACTCCCACTTACTGACCTCTCTGTCCCTG	25522
Oy	2643	-----SerTyrAsnCysProSerProTyrAlaAspIleArgLeuTyrGlnProTyrGlnYA	2661
Db	25523	CTTAGTCTTAACAATCGGCCAGGCCCTTAAGCAGATGGGGGGGTGTACAGCCTTAGGGAG	25582
Oy	2661	epsSerAlaGlySerLeuHisSerThrSerXysSerGlyIlybSerGlnProSerTyrIleP	2681
Db	25583	ACTGGCGCGGTCTCTGTCACAGCACCAAGTGTCTGGGCAAGATCAGCCACGTAATCATCC	25642
Oy	2681	roPheLeuLeu-----	2684
Db	25643	CCTTCTTGCTAGAGTGAATCCCGGAGATGGGAAGGTGGAGAGGGAGAGGGGCCACAG	25702
Oy	2684	-----	2684
Db	25703	CATGCTGAGACCCAGGCCAGCGACGCTGTGGAGTTGAGAGACACACTGTGGTGAAGT	25762
Oy	2684	-----	2684
Db	25763	GGGGGCCAGCTTGAATTGAAGCTGTAAAGGACCCACAGCAGGACCAAGAAATCCAGGGG	25822
Oy	2684	-----	2684
Db	25823	AGAGAGAGACCTGGGACCTGGGCAAGGGGCCAGGCTGACCCCTCCAGCATGTGTCTCATC	25882
Oy	2685	-----ArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyIleLeuGlyAspProG	2703

Db	25893	TTCTTAAGGAGAGATCCGACCTGAACCTTCGACCAAGGGCCCTCGGCTGTGGAGATCCAG	25942
QY	2703	lySerLeuPheLeuGluGlyGlnAspGlnGlnHisAsp-----	2715
Db	25943	GCAGCCGTTCCTTGAAAGGTCAAGAACAGAGCATGG-TAAGGACAGAAAGCTCTGGCCA	26001
QY	2715	-----	2715
Db	26002	CCACGAGGGCAGATTGGCTGGCTTTTACTGAAGTGGGTGGAGGTGGCTGGCTGT	26061
QY	2716	-----ProAspThrAspSerAspSerAspLeuSerLeuGlnA	2728
Db	26062	GATCTCTCCCTGGGCTCTTAGATCTCGACACGGACTCCGACAGTGACTGTCTCTTAAGAG	26121
QY	2728	spAspGlnSerGlySerThrAlaSerThrHisSerSerAspSerGluGluGluGluGlu	2748
Db	26122	ACGACACAGAGTGGCTCTTATAGCTCTTACCACCTCAACACAGTGAAGGAGGAAGAGG	26181
QY	2748	lunGluGluGluGluAlaAlaPheProGluGluGlnGlyTrpAspSerLeuLeuGlyProG	2768
Db	26182	AGAGGAAAGAGAGGGCGCGCTTCCCTGGAGAGAGAGGCTGGATAGCTGTGGGGCTTG	26241
QY	2768	lyAlaGluAlaGluPheLeuHisSerThrProLys-----	2779
Db	26242	GAGAGAGAGACGTGGCCCTTGCAACAGTACCCAAAGGTGGGCGCACACTGGGCTGTGGC	26301
QY	2779	-----	2779
Db	26302	CTTTGGGGCCAGTGGGAGAGACAGTGGGCTGGGGTTCTTGGAGACAGACTGGGCTGT	26361
QY	2780	-----AspGlyG	2782
Db	26362	GGCCTCGGTGATCTGACCCCTGTGGGGCTCATCTACTTCTTCCACCAAGATGGGG	26421
QY	2782	lyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlnSerS	2802
Db	26422	GCCCAAGGCTCTGCAAGGCCCCCTGGCCAGAGACTTTGGGACCAACAGAAAGAGTA	26481
QY	2802	etGlyAsnGlyAlaProGluGluAlaGlyLeuAlaGlyLysGlyLysAlaLeuSerArgGluG	2822
Db	26482	GTGGCAACGGGGGCCCCGTGAAGAGCGGCTGCGGAGAAATGAGATGCCCTGTCTTGAGAGG	26541
QY	2822	lySerLeuGlyProLeuProGlySerSerAlaGlnProHisLys-----	2836
Db	26542	GGTCCCTAAGCCCCCTTCCAGGCTCTTCTGCCCAAGCTTCAAAAGTGATGGGGCACCC	26601
QY	2836	-----	2836
Db	26602	CCAGCTGCGAGAGCTCCCTAAGTCAAGAGGCTCATACCTCACTTCTCTGTGGCCGCAC	26661
QY	2837	-----GlyIleLeuLysLysGlyCysLeuProThrIleSerG	2849
Db	26662	TCAAGACCCCGCCCCCGGCCACACAGGAGTCTTAAAGAAAGTGTGTGCCCAACATCAAGG	26721
QY	2849	lyIysSerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSerArgLysSerS	2869
Db	26722	AGAAAGACAGCCCTCTGCGGCTTCCCTTGAGACAAATGACAGAGGCTTCTCGGGGCTCTCT	26781
QY	2869	exAlaSerGluGlySerArgGlyGlyProProProArgProProProArgGlnSerLeuG	2889
Db	26782	CCGCTGTAGAGGCAAGCGGGAGGGCCCCCTCCCGGCCAACCGCCCCGGGAGAGCTGCC	26841
QY	2889	lunGluGluLeuAsnGlyValMetProIleAlaMetSerTlleValAlaGlyThrValAspG	2909
Db	26842	AGGAGAGCTGAAACGGGGTCAATGCCATTCATGAGATCAAGGCAAGGCAACGGGTGATG	26901
QY	2909	lunAspSerSerGlySerGln2915	
Db	26902	AGGACTGTCTAGGCTCCAGG26921	

DEFINITION Homo sapiens chromosome 1 clone RP11-6L21, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION BX284647

VERSION BX284647.1 GI:28865245

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 227194)

AUTHORS Hall, R.

TITLE Direct Submission

JOURNAL Submitted (03-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT ----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: b66121

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ABI; 0% of reads

Chemistry: Dye-terminator ABI; 96% of reads

Chemistry: Dye-terminator Big Dye; 3% of reads

Consensus quality: 16459 bases at least Q40

Consensus quality: 16460 bases at least Q30

Consensus quality: 16463 bases at least Q20

Insert size: 227194; sum-of-ctigs

Insert size: 168516; 3.2% error; agarose-fp

Quality coverage: 7.72x in Q20 bases; sum-of-ctigs Quality coverage: 10.52x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 227194: contig of 227194 bp in length.

FEATURES

source

1..227194

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="1"

/clone\_idb="RP11-6L21"

1..227194

/note="assembly\_fragment:02685

clone\_end:SP6

vector\_side:right

clone\_end:T7

vector\_side:left"

misc\_feature

1..227194

/note="assembly\_fragment:02685

clone\_end:SP6

vector\_side:right

clone\_end:T7

vector\_side:left"

BASE COUNT 64124 a 52119 c 51230 g 59721 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 227194

Score: 12339.00 Matches: 22907

Percent Similarity: 36.92% Conservatives: 0

Best Local Similarity: 36.92% Mismatches: 8

Query Match: 79.38% Gaps: 4967

DB: 2 Gaps: 32

US-09-916-849a-3 (1-2923) x BX284647 (1-227194)

QY 1 MetargerProAlaThrGlyValProleupProThrProProProleuleuLeu 20

DB 3330 ATGGGAAACCCGGCCACCGGGGTCCCTCCCAAGCCGCGCCGCTGTGTGCTG 3389

QY 21 LeuleuleuleuLeuProProProleuleuGlyAapGlnValGlyProCyAargSerleu 40

DB 3330 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449

QY 41 GlySerAargGlyAargGlySerSerGlyValAcyAalPrometGlyTProleuCyProSer 60

DB 3450 GGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3509

QY 61 SerAlaSerAamLeuThrLeuThrLeuThrSerAargCyAargAapAaglyThrGluLeuThr 80

DB 3510 TCAAGGTGGAACCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 3569

QY 81 GlyHisLeuValProHisHisAapGlyLeuAargValTProProGluSerGluHis 100

DB 3570 GGCCACCTGGTAACCCACACAGATGGCTGAGGGTTTGCTCCAGAAATCCGAGGCCAT 3629

QY 101 IleProleupProProAlaProGluGlyCyAProTProSerCyAargLeuGlyTleGly 120

DB 3630 ATTCCTTACCAACAGCTCTGAGAGGCTGCTGAGCTGTGCTGCTGCTGCTGCTGCT 3689

QY 121 GlyHisLeuSerProGluGlyValLeuThrLeuProGluGluHisProCyAleuValA 140

DB 3690 GGCCACCTTCCCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3749

QY 141 ProAglLeuAargCyAargGlnSerCyAglLeuValAglAalAProGlyValAargVal 160

DB 3750 CCACGCTCAATGCAATGCAATGCTGCAAGCTGCAAGGCTGCAAGGCTGCAAGGCTG 3809

QY 161 ArgSerProGluGluSerLeuGlyValAargGlyValAargAanValAamThrAlaProGlu 180

DB 3810 AGGTACCAAGAAAGAGTCCCTGGGTGGCTGGAGAAAGAAATGAAATACACCCCCAG 3869

QY 181 PheGlnProProSerGlyGlnAlaThrValProGluAengInProAlaGlyThrProVal 200

DB 3870 TTCCAGCCCCCAAGTACCAAGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3929

QY 201 AlaSerLeuAargAalAaAapProAapGluGlyGluAglAargGluGlyThrMet 220

DB 3930 GATCTCTGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 3989

QY 221 AapAlaLeuAapAapSerAargSerAamGlnPhePheSerLeuAapProValThrGlyA 240

DB 3990 GATGCTCTTGTATAGCCGCTCCCAACAGTCTTCTCCGAGACCAAGTCACTGTG 4049

QY 241 ValThrThrAlaGluGluLeuAapAargGluThrLySerThrHisValPheAargValThr 260

DB 4050 GTAAACCAAGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4109

QY 261 AlaGlnAapHisAglMetProAargAargSerAlaAamAlaThrLeuThrHisValThr 280

DB 4110 GGGCAGGACCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 4169

QY 281 AapThrAapAapHisAapProValPheGluGlnGlnGlnGlnGlnGlnGlnGlnGln 300

DB 4170 GACACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCA 4229

QY 301 AamLeuGluValGlyTyrGluValLeuThrValAargAlaThrAapGlyAapAlaProPro 320

DB 4230 AACCTGAGGAGTGGCTATGAGGAGTGGCTATGAGGAGTGGCTATGAGGAGTGGCT 4289

QY 321 AapAlaAanIleLeuTyrAargLeuLeuGluGlySerGlyGlySerProSerGluValPhe 340

DB 4290 AATGCCAATATTCCTGACCGCTGCTGAGAGGGGCTGAGGGGAGCCCTTCAAGTCTTT 4349

QY 341 GlnIleAapProAargSerGlyValIleAargThrAargGlyProValAapAargGluVal 360

DB 4350 GAGATGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4409

QY 361 GluSerTyrGlnLeuThrValGluAlaSerAapGlnGlnAargAapProGlyProAargSer 380

```

Db      4410 GAATCTTACGAGTGAAGGTTAGGACAGAGTCGAGGACCGGGCTCTCCGAGT 4469
Qy      381  ThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAspAlaProGlnPheSer 400
Db      4470 ACCACAGCGCGCTGTTCTCTTCTGTGAGATGACAAATGATATCCCGCCAGTTTAAAT 4529
Qy      401  GluLysArgTyrValValGlnValArgGluAspValThrProGluValaProValLeuArg 420
Db      4530 GGAAGCGCTATGTGCTCAGGTGAGGAGATGTACTCCAGGGGCCCGCAGTACCTCGA 4589
Qy      421  ValThrAlaSerAspArgAspLysGlySerAspAlaValValHisTyrSerIleMetSer 440
Db      4590 GTACACGCTCGGATGAGACCAAGGAGACCAATGCGGTGTCATATGACATCAAGAT 4649
Qy      441  GlyAspAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAspValValSer 460
Db      4650 GGCAATGCTCGGGAGACAGTTTATCTGAGATGCCCAAGCTGAGACTTGAATGTGTGAC 4709
Qy      461  ProLeuAspTyrGluThrThrTyrGlyTyrThrLeuArgValArgAlaGlnAspGlyGly 480
Db      4710 CCTCTTGAATATGAGACCAAGAGTACACCTTACGGGTGACACAGATGTGTGAC 4769
Qy      481  ArgProProLeuSerSerValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500
Db      4770 CGTCCCGCACTCTAATGTCTCTGCTGTGAGACAGTACAGGTCTGTGATATCAAGAC 4829
Qy      501  AspAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db      4830 AATGCCCATCTTCTCTCAGACACCTTTCAGAGCTACTGTCTTGAGAGTGTCCCTTA 4889
Qy      521  GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeu 540
Db      4890 GGTACTCTGCTCTCATGTCTCAGGCTATCGACGCTGATGTGTGTGACAAATCCCGCTG 4949
Qy      541  GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAspAsnGlyTyrGly 560
Db      4950 GAATACCGCTGTGCGGTGAGACATGACTTCCCTTCAACATCAACATGAGCACAGGC 5009
Qy      561  TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db      5010 TGGATCTGTGTGGCTGTAATGAGACCGGAGGAAGTTGATTTCTACAGCTTTGGGGA 5069
Qy      581  GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db      5070 GAAGTCTCGAGACCAATGCGCACTCAGACACTCAGCTCGGCGGACAGTGTGAGCTGCTG 5129
Qy      601  LeuAspValAspAspAspAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db      5130 CTGAGTTCACGACCAACATCAACCTTTACCAACCAAGATACACAGTGGGCTCAAT 5189
Qy      621  GluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db      5190 GAGAGTGAAGCTGTGGGACCAAGCGTGTGACGCTGTGACCGTGTACTCTCAT 5249
Qy      641  SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAspAsnArgPheSerIleThrSer 660
Db      5250 AGTGCATCACTACCAATCAAGTCAACAGTGGCAATATCTCGAACCCTTCTCCATCAACGC 5309
Qy      661  GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArgGln 680
Db      5310 CAAGATGGTGGGTGGCTGTATCCCTGTGCTGCCATCGAGCTACCAAACTTGAAGGGGAG 5369
Qy      681  TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db      5370 TATGTGTTGGCTTACCGCTCCGATGGCACTCGGAGAGACCGGCAAGATTGTGGTG 5429
Qy      701  AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db      5430 AATGTACCGACGCGCAACCACTCATGTCTGTCTTCAAGAGCTCCACATATCAAGTGAAT 5489
Qy      721  ValaGlnGluAspArgProAlaGlyThrThrValValaLeuIleSerAlaThrAspGluAsp 740

```

---

```

Db      5490 GTTAAATGAGACCGGCGCGGACGACCAAGGTGTGTGATGACGCGCACGAGTAGAGAC 5549
Qy      741  ThrGlyGluAspAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db      5550 AAGAGTGAAGATGCCCGCATCACTTCAATGAGAGACAGATCCCGCAGTTCCGCAATC 5609
Qy      761  AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db      5610 GATGCAGACAGGGGGCTGTACCAACCAAGGTGAGCTGACATTAAGAAGACCAAGTCTT 5669
Qy      781  TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrTyr 800
Db      5670 TACACCTGTGGCATTAATGCTCGGAGCAATGGCATTTCCCGAAGATCCGACACCACTAC 5729
Qy      801  LeuGluIleLeuValAspAspValAspAspAspAlaProGlnPheLeuArgAspSerTyr 820
Db      5730 CTGAGATCTGTGTAGACAGTGAATACAAATGCCCTCAGTTCTGTGAGATCTCTAC 5789
Qy      821  GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db      5790 CAGGCGAGTGTCTATGAGAGATGTGCACCTTCACTAAGCTCTGCAATCTCAAGCACT 5849
Qy      841  AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860
Db      5850 GATCGTGAATCTGCACTTAATGACGAGGCTCTTCAACCTTCAAGAGGCGCAGATGGA 5909
Qy      861  AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLeuAsp 880
Db      5910 GACGTGACTTATTTGTGATGTCACGTCAAGCAATGTGCGAACCTTACGGAGGCTGAAT 5969
Qy      881  ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValaAspLysGlyMetProPro 900
Db      5970 CGAGAGACGTGGCCCAAGTATGTCTTGCGGCATATGACAGTGAACAGAGGATGCCCA 6029
Qy      901  AlaArgThrProMetGluValThrValThrValLeuAspValAspAspAspProVal 920
Db      6030 GCCCGCACCACTTAAGAGATCAAGTCACTGTGTGATGTGAATGACAAATCCCTCTGTC 6089
Qy      921  PheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db      6090 TTTAGAGAGATGAGTTGATGTGTGTGTGAGAGAGAACACCCCATTTGGGCTAGCCGTG 6149
Qy      941  AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db      6150 GCCCGGTCACAGCACCTGACCCCGATGAAGCACCAATGCCAGATTATGACCAAGATT 6209
Qy      961  ValGluGluAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla 980
Db      6210 GTGAGGGCAACATCCCTGAGAGTCTTTCAGCTGACATCTTCTCCGGGAGCTGACAGCC 6269
Qy      981  LeuValaAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThrSer 1000
Db      6270 CTGTGTAGACTTAGCTACGAGACCGGCTGAGTACGTCTGTGTCAACGCGCACTGCA 6329
Qy      1001  AlaProLeuValSerArgAlaThrValHisValaArgLeuLeuAspArgAspAspPro 1020
Db      6330 GCTCCTCTGTGAGACCGGGCTACAGTCCAGTCCGCTTGAACCGCAATGACCAACCA 6389
Qy      1021  ProValLeuGluAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040
Db      6390 CCAATGTCTGGCAACTTGAATCTTTTCAACAACATATGTCAACAAATCCCTCAAGACGC 6449
Qy      1041  PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db      6450 TTCCCTGGGGGTGCATTGGCCAGATACCTGCCATATCTCAGATATGATCTG 6509
Qy      1061  ThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db      6510 ACTTACAGCTTTGAGCGGGGAATGAATCACTGCTGTCTGCTCATATGCTCCACGGGT 6569
Qy      1081  GluLeuLysLeuSerArgAlaLeuAspAspAspArgProLeuGluAlaIleMetSerVal 1100
Db      6570 GAGCTGAAGCTAAGCCGCGCATGAGCAACACCGGCTCTGAGAGGCCATCATATGAGCGTG 6629

```

QY 1101 LeuValSer----- 1103  
 Db 6630 CTGGTGTTC-AGGTAAAGAAAGGCCAGATGCGCTGGGGTGGGGGTAGTCTCGCGGGGATG 6688  
 QY 1103 ----- 1103  
 Db 6689 GTCTGGGACACCACTGAGGTAGGGTGCATCCAGAAAGCACTACAGATCCACTCCCT 6748  
 QY 1103 ----- 1103  
 Db 6749 GCCCAGTGCCTGGCAGAGACAGAGAGGGGCAAGGCCAGCTTGGAAAGAACCCCAAGAA 6808  
 QY 1103 ----- 1103  
 Db 6809 CCTGGCGGCTGGTGCAGGCCCCACTTCTGTGGCTGGGCTGGGCTGTGGTTGGC 6868  
 QY 1103 ----- 1103  
 Db 6869 TGGCCCCACTGTCTACCAAGCTGTGAAGTGTGGGAAAGTGTGTGAAGGCTGCCCCGAC 6928  
 QY 1103 ----- 1103  
 Db 6929 TATGCTGCCAAGATTGTGAAAAAGGGCGTGAAGCTGGGGGGCCAGAGAGCTGG 6988  
 QY 1103 ----- 1103  
 Db 6989 GGGAAAGGAGAGAGGACTGTGGCGCAGAGAGAGGCGAGGCCAGGGCCCTTCAACCC 7048  
 QY 1103 ----- 1103  
 Db 7049 CATCAAGATATTTCTGCAAGTCTCAACAAGCTTTGGCAAGACTGGGGAACACGCGAGCAC 7108  
 QY 1103 ----- 1103  
 Db 7109 AGGGCTGCGGCAACCCAGGGGTGAGAACCTGTAGTCTCTTGGCCCCCTTTGGCTCC 7168  
 QY 1103 ----- 1103  
 Db 7169 TGGCTGAGCTGCAGAAATCGGCTGACAGGACTTGTGTACGCGATTCTTTCCTCCAGAG 7228  
 QY 1103 ----- 1103  
 Db 7229 CAGAAAGAGCTGGCTGGCTCCCCCTTCAGAGCTCCACAGGAATGAGGCCAGAGCTC 7288  
 QY 1103 ----- 1103  
 Db 7289 ATTGTCTGTCAAAACAGACCCCACTTTCAGAGTCCGTTGGTCTGGGTTGGAGAG 7348  
 QY 1103 ----- 1103  
 Db 7349 GGGCTGCATTAAGAGCTGGGGGGGAGGCTGAGGCTGTGGGCCCCCTTGCACCTCATG 7408  
 QY 1103 ----- 1103  
 Db 7409 GGGTGTCTTAGGGTGCAGGCTTCTTCTGTAGCTGTGACCAAGCCCAAGCTCCAG 7468  
 QY 1103 ----- 1103  
 Db 7469 GAATGAGGTAGGAGACAGACAGGCAAGTGAAGTGTGAGGCTTGAAGGACAGAGACC 7528  
 QY 1103 ----- 1103  
 Db 7529 CAAAGCACTTATCTAGAGAGAGACAGAGTCCCAAGGCGTGGAGCTGGCCGGAACCC 7588  
 QY 1103 ----- 1103  
 Db 7589 AGGCTGGGGGCAATTCGTGCTGTGGCAAGCACTCAGCTCAGACTCCCTGAGACAGAG 7648  
 QY 1103 ----- 1103  
 Db 7649 GGCTGGGGAGAGAGTGTGAGGGGGTGTGTGACAGTATTTCTGTCTTTTATTCAGCTG 7708

QY 1103 ----- 1103  
 Db 7709 TGGGAGCCAGAGTAGAGAGAGGAGAAACACACCAAGTTGTGAGTCTTCCCC 7768  
 QY 1103 ----- 1103  
 Db 7769 AAGCTGGAAACACTGAGACCTCATAGAGGGATGTGTAGCTGACCCCAAGGGGTGG 7828  
 QY 1103 ----- 1103  
 Db 7829 ATCAGGTCAATCCCTCTTAGGACGCTTCAATTCATCAGGGAGGGAGAAAGAGA 7888  
 QY 1103 ----- 1103  
 Db 7889 GGCTGGAGGCCCTTATCCACCCAGAGAGTGTGCTTGAAGGGAGAGAGGGTATG 7948  
 QY 1103 ----- 1103  
 Db 7949 CCAAGTCAAGGCGTGGCTCACTCAGGGTCTTGTGACAGATGCTCCCTGCCAC 8008  
 QY 1103 ----- 1103  
 Db 8009 CCTCAAGTAGCAACAGGAACAGAAAGAGAGCTTTTCAAGGACTGTGGCTAGGTT 8068  
 QY 1103 ----- 1103  
 Db 8069 CTGAGCTTCTGAGACTTCAAGTTTCCCTCCAGAGGCGGAAGAGGCTGAGCTCC 8128  
 QY 1103 ----- 1103  
 Db 8129 TGTGTCTCAACTGTGCTGGGCACTTGAACAAAGTGGAAAGAGGTGGAAAGAGGCC 8188  
 QY 1103 ----- 1103  
 Db 8189 TCTCCCTGTGCTCTCAACACCAAGCTGCTTGTGTAGCGTGGCAGGCACTGGCAGA 8248  
 QY 1103 ----- 1103  
 Db 8249 GAAAGCCCCATTAAAGCTTGGTGTCCACAGAGGAGAGCCCTTAGACTTGTGCC 8308  
 QY 1103 ----- 1103  
 Db 8309 TAGTCTAAGAAAGGCTCGGGGGAACATCCAGAGACGGCTCTCTGGCCCCGAGCTTGG 8368  
 QY 1103 ----- 1103  
 Db 8369 CCCCACTTCTCCCTTGGGCTTGAACCTTGGAGGCACTCACAGGCAACCCAAAGTT 8428  
 QY 1103 ----- 1103  
 Db 8429 CAGGAGGCCGGGTGAGGGGAGGAGTGACTCCACTGCCCCCTTGTCTTCTCTGTGT 8488  
 QY 1103 ----- 1103  
 Db 8489 TCTTGTCTCTTCTTCTTCCGCCCCCTGCCCAAGGGAGCCTCAGACATGCTCTC 8548  
 QY 1103 ----- 1103  
 Db 8549 TCCGGTGCCTCTCCCTGCTCTGCGGCTCTCCGCTCTCCACTTGGCCATTTGTGGCT 8608  
 QY 1103 ----- 1103  
 Db 8609 CCTTCTGAGGCTTGTGTCTTCCCTGTTTGTGTCTCCGTTGTGCTCCACTCTCTTC 8668  
 QY 1103 ----- 1103  
 Db 8669 CCTATGGCTGAGCTGCCCAAGCTGGCCGATGGCTCCCTGAGGCTGAATCTCTCTC 8728  
 QY 1103 ----- 1103  
 Db 8729 GGCTCTTCTTCTCTGTCTCAGACAGCAGACTCTGTGTCAAGTGTCTGGCAGCTCCC 8788  
 QY 1103 ----- 1103

Db	8789	AGAACAGACAGACCCTCTTGGCCTCCAGTCTCCGCTGGCACAATCGCAGGGGAGAG	8848
Qy	1103	-----	1103
Db	8849	GCTGGAGGCTGACCGGGAACCTGGCACTTTTCAGGTGTCCCTGAGCTTAGCTTGGACAC	8908
Qy	1103	-----	1103
Db	8909	AGTGGCATTTCTGGGGTCAGAGGCCATCCAGCTGTTCTTTCTGTCCAGGAAGCA	8968
Qy	1103	-----	1103
Db	8969	GCTCAGGCAACGTACTCCCAACCCCTGTTCATCTGCTTGTGCACTGAGGGGTGACC	9028
Qy	1103	-----	1103
Db	9029	TAGGTCMAAGCAGGGGAGGGTGATATAGAAAGAGAGACAGTGGGTTGACTGTTGGGA	9088
Qy	1103	-----	1103
Db	9089	TGGGGACAACAGCAGAGAGGGGCAAGTGGCCAAGATGTGGACCAAGATGGGAAGCAACA	9148
Qy	1103	-----	1103
Db	9149	GAGCATGTGAGAGAGGGGCCAGGCGGGCTTAAGACGATGGCGGTGGGGGACACAGG	9208
Qy	1103	-----	1103
Db	9209	GAGAAGTGTGTTGTTCCAGGGCTGAGGGGTGGCAAAATTCTCCCGCTCAATGGCCTTGG	9268
Qy	1103	-----	1103
Db	9269	TTGCCCTCTCTCTTTAATCAGTTATGAGGCACTGTGTGAGATGTGGCTTGGCTT	9328
Qy	1103	-----	1103
Db	9329	CTGCTTGAGGGAATATCCCGTCAAGAAATCATTAGCGATATATACAGGTATACAAG	9388
Qy	1103	-----	1103
Db	9389	CTGTGTCCAGCTAGCACTCTGTGGGGGGAAGTCTGAAAACGAGAGAAAGTTAGACC	9448
Qy	1103	-----	1103
Db	9449	TGATGATATAATCTGGGCAAGCTTCTTGAGCAGTGAAGTAGGCAAGTTAGAACT	9508
Qy	1103	-----	1103
Db	9509	CAGAGACTGTGGCTGTGTCTGGTGGAGAGAACAGCTGTGACAGACTTTGAGAC	9568
Qy	1103	-----	1103
Db	9569	GAGAGCAGAGGAACAGAAAGTGGGTGATGACAGACAGGCTTGTGTGAGAGCTTAGAG	9628
Qy	1103	-----	1103
Db	9629	AATTAACAAGTATGAGGCTGAGGGTAGTGGSCAGTCCGCTGAGGACATGGCTGAGAT	9688
Qy	1103	-----	1103
Db	9689	GGGGGTCTGCTTTCAGCCTTCCCGCTTCAGCTGAGCAACGCTGTACAGTCAAGCCACC	9748
Qy	1103	-----	1103
Db	9749	GCAAGTAGTACACACAGATGCAATTTGTTCCATGTCCCTTGCTACTCCCACCC	9808
Qy	1103	-----	1103
Db	9809	ACCCAGAAAACATGGCTCTGCCCCGTACTGCTAGGGTAAAGCAATTGCTTTTATATC	9868
Qy	1103	-----	1103

---

Db	9869	TAGTGGGCTCCCCCTGTATGTGTGTTGTTGTTGTTGGACAGAGTCTCACTCTG	9928
Qy	1103	-----	1103
Db	9929	TCAACCAAGCTGAGTACAAATGGCGCATCTCGCTCACTGCAACCTCTGCTTGGGT	9988
Qy	1103	-----	1103
Db	9989	TCAAGCAATTCCTGCTCAGGCTCCGTAGTATCTGGGACTTCCTGATAGTGAATT	10048
Qy	1103	-----	1103
Db	10049	ACAAGCAACCAACACAGCCCACTAATTTTGTATTTTATAGAGACGGGTTTCAC	10108
Qy	1103	-----	1103
Db	10109	CATGTGTGATGTGATCTGAACTCTGACCTCGTATCAACCGGCTCAGCCTCCA	10168
Qy	1103	-----	1103
Db	10169	AAGTCTGGACTACAGGGGTGAGCACTGGGCCCGGCCCTGTATGTTTAAAGTCC	10228
Qy	1103	-----	1103
Db	10229	GTTTATCTTGTGAGACTGAATAGATGTATCACTCACTCAGTGTTCAGAACCC	10288
Qy	1103	-----	1103
Db	10289	TACTATGTGAAGCATTTGTAATAATAACAATATGAATATGCCATAGCTCATGCTGT	10348
Qy	1103	-----	1103
Db	10349	AATCCAGTCTTAGAAGCTGAGATGAGATCACTTGAGGCCAGAAATTGAGACCA	10408
Qy	1103	-----	1103
Db	10409	GCCTGACACACAGCAAGACCCCATCTCTACACACAAATAATAGCCAGGCATATG	10468
Qy	1103	-----	1103
Db	10469	TGTAGCACTGTGTGTACGCTAATTGGAGGTTGAGGTGAGAGATCCCTTAGCCCA	10528
Qy	1103	-----	1103
Db	10529	GAAGTCAGAGCTACAGTAGATATGATGTGCCATGTCTTCAAGTATAGGACAGAG	10588
Qy	1103	-----	1103
Db	10589	AGTCTGTCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	10648
Qy	1103	-----	1103
Db	10649	CTCTCTTCCAGGCCCAAGCTTCAATTTTATAGTAAATAAATAATTCATTCATGTGTA	10708
Qy	1103	-----	1103
Db	10709	TGTCATTCAGGTGTATATATTTCTGGGCTTAGGTTGTGTGAGTAATGACACCC	10768
Qy	1103	-----	1103
Db	10769	AGCAAGTGTATTTAGTCTTAGCTGATTTATATAGTAGTGTGTGCTGGGGCATAGCA	10828
Qy	1103	-----	1103
Db	10829	GGCATTTGGGTGTAATGTGTGAGCCAAAGTGGGTATGATAGCTAGAGGCCCTAGAACCA	10888
Qy	1103	-----	1103
Db	10889	GAGAGAGACTGACTGAGGGGCTCTGGGGTCTCAGAGTGAAGACTCCAGGCCCTT	10948
Qy	1103	-----	1103
Db	10949	CTATCTGTGCTTCCGGTCCCAACCCGACTCAACCTGTCTTCAGAGATGGGGAGTGGG	11008

QY 1103 ----- 1103  
 Db 11009 TGAATGATCACTGTGTTAAAGTGGCTCAGCTGATTTCTGAGGGTGGGGCCCCCAGGGCT 11068  
 QY 1103 ----- 1103  
 Db 11069 ACCCAACAGACCCCAAGAGAGTGGGGGGAAGATGCTTTGTGTGGATGTTGCCATGGG 11128  
 QY 1103 ----- 1103  
 Db 11129 GATAGCAGGCTGTGGGGCCCAAGACAGCATGGGGCTCAGTGGGAGGGGAGCTAGTGGCC 11188  
 QY 1103 ----- 1103  
 Db 11189 CACTGACCCAGTTTACCACTCCCCCAACCCCGCAACTCCCTGTGCTGTGGGGGAG 11248  
 QY 1103 ----- 1103  
 Db 11249 TGGGGAACACAGCCATGCTTGCATTAATGCTGTGTGAGGTGTAGAGGCTGAGGGGGCA 11308  
 QY 1103 ----- 1103  
 Db 11309 GGGGATGATTTTGTGGGTCTGAGGATGCTCTGGGCCCTGTAGAAAAGTATGAAAAAGA 11368  
 QY 1103 ----- 1103  
 Db 11369 CAACATAGGGTTCGCCGAGAGGAGCAGACACTTGGGGGAGAGGAGCAGAGCGGTGGC 11428  
 QY 1103 ----- 1103  
 Db 11429 TTGGGCAAGGAAAGGTGCACAAACATTAGAGGTGTGCGTGCATCAGCAGCAAGAGTT 11488  
 QY 1103 ----- 1103  
 Db 11489 TTGGGAAGAGATGCAAAGAGACTGCAAATCTCATCCCTGAGCCCTGAAAGAGTATTG 11548  
 QY 1103 ----- 1103  
 Db 11549 GCCCACAAGTGAATCCTGGCTCAGCACCTTATGCAAGTTGAAAAAGAGACCTTTCT 11608  
 QY 1103 ----- 1103  
 Db 11609 TGGTGGGTGTGACAGAAATTCAGAGCTGAATTCAGAGCCCAAGGCTGGCTCTGACTGTG 11668  
 QY 1104 ----- 1104  
 Db 11669 TCCCTCTCCACAGACGCGCTACACAGCGTGAACGCGCAAGTGGCGCTGCGTGAACATC 11728  
 QY 1120 TTTTThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 Db 11729 ATCACCAGTGAATGCTCAGCCACAGCATCAGCTGGCCTGAGGAGCAATGTCACCCGAG 11788  
 QY 1140 ArgPheLeuSerProLeuLeuGluPheIleGluAlaValAlaIleThrLeuAlaThr 1159  
 Db 11789 CGCTTCTGTACACACTGCTAGGCTCTTTCATTCAGGCGGTGGCCGCAAGCTGGGCAAG 11848  
 QY 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
 Db 11849 CCACCCGAGCAGCGTGGTGTCTTCAAGTACAGCGGAGCACCGACCCCGGGGGGCAAC 11908  
 QY 1180 TTTTLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPhe 1199  
 Db 11909 ATCTCAACCTGAGGCTGTGTGGTGGCCAGCCGCGAGGGCCGCGGGGCGGCGCTTTC 11968  
 QY 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 Db 11969 CTGCGCTCTGAGGACCTGCAAGAGCGGCTTATCTTCAACCGCAGCTGTGACGGCAATC 12028  
 QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
 Db 12029 TCGGCAACAGCGGTGCTGCTTTCAGACAAACATCTGCTGGGGAGCGCTGCGAAGAAC 12088

QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
 Db 12089 TACATGCGCTGTGGGTGTGGGTGTGGCTTCCAGCTCTCCGCGCCCTTATGCTCTCTCC 12148  
 QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
 Db 12149 TCCGTGCTCTTCCGGCCCAATCCACCCGCGAGGGGCTGGCGTGGCCGCGCGCGCGG 12208  
 QY 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
 Db 12209 TTCAGGGGTGACTACTGCGAGACCCAGGGTGAACCTCTGCTACTGCGCGCCCTGTGGCC 12268  
 QY 1300 HisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
 Db 12269 CAGGGCGCTGCGCAGCGCGAGGGCGGCTTACCTGCTGTGTGATGGCTTACAG 12328  
 QY 1319 ----- 1319  
 Db 12329 GGTGAGCCAGGGAGGGAGCTTCATGAGGCCAGCCCTGGAAGGCTGACTGTGTGCAAGC 12388  
 QY 1319 ----- 1319  
 Db 12389 ACAATCAGACAAATGCTGGCGGCTCATTTCTTCCGAGTGAAGTGCACCTACA 12448  
 QY 1319 ----- 1319  
 Db 12449 CTGAGAGTCAATAATCCGGCTGCTCCCTGACAGACCCCACTGAGGAGGAGAGTGCAG 12508  
 QY 1319 ----- 1319  
 Db 12509 GAGTGTCTTCTCCGTGTGTGACAGACACAGCAGAGTGCATTTGGGCGAGACTGG 12568  
 QY 1319 ----- 1319  
 Db 12569 CTCCAGACAGTGGCGCTGTGTGTGATCAGCTTGGCGCTAGTAGCTGTGACCT 12628  
 QY 1319 ----- 1319  
 Db 12629 TGAACAATTAATTAACTTTCGTGACCTCGGTTATTGAAATGAGATATAGACCTTAC 12688  
 QY 1319 ----- 1319  
 Db 12689 TTAGGTCAATAGTGTGTGTGAGATTATATTAATTGATCTTGGAAAAAGCTGG 12748  
 QY 1319 ----- 1319  
 Db 12749 TGGCATGGAAGGTGTGGCTTGAATTTAAGGGAAGCTGAGATTGGGTGACTTGTCTGG 12808  
 QY 1319 ----- 1319  
 Db 12809 GTCTCATCACACCGGTTTAGAGGTGAGGCTGGGCTTAGAACCTGTGTCTCCCCACTGG 12868  
 QY 1319 ----- 1319  
 Db 12869 AGCGAGTTCCTTTCTGAGCTGACAGCAAGTACTTTCTCCAGGCAATTAAAGGGAAGAC 12928  
 QY 1319 ----- 1319  
 Db 12929 ACAGAGAGCTGGGGGAATGGAAGGGGTGAGTTCTGAGATGGGGTCTCCAGAAACC 12988  
 QY 1319 ----- 1319  
 Db 12989 ATTGCACTGTATGGGCTGTGGCAGATAGCTCGGCTTCGCGCCAGCTGTGTAAAGA 13048  
 QY 1319 ----- 1319  
 Db 13049 GGATCTGTGAGAGCGGCTTCCCTGTGGCTGTGACGCAAGTCTTTAACATCTGTCT 13108  
 QY 1319 ----- 1319  
 Db 13109 CCTGGAGTGCATCTGTCTGTCTATATGCTCTGACCCCTTCCCTTCAAGGCTCTGTGT 13168  
 QY 1319 ----- 1319

Db 13169 CTTCTCAGAGACACCCCTGCTCCCTTCCCTGTGGCCCTTCTCTAGATGCCAGGC 13228  
 QY 1319 ----- 1319  
 Db 13229 CCTGCTCTTGTCTTCAGAGGCCATGCCACTGCTCTGCCAGACCCATTAGCAGGCC 13288  
 QY 1319 ----- 1319  
 Db 13289 CACTCTTCCCGCTGCATCTTCTGTCCAGAACTTGCAACTTGAATTCTCTGTGCTG 13348  
 QY 1319 ----- 1319  
 Db 13349 CTGCTAGGGGTGGGGGAAAGGGAGGACGGGCTGGGGGAGGAGAAAGTCTTGTCTC 13408  
 QY 1319 ----- 1319  
 Db 13409 CAGAGGATGTGTGGGGGGGTGGAGAGGGGGGGTTAGACAGCGCAACACAGTTTCC 13468  
 QY 1319 ----- 1319  
 Db 13469 ATGAAACAGAGCCAGTCACTAGGCAAGAGCTGAGAACTTATCTTAGAATCTGGGA 13528  
 QY 1319 ----- 1319  
 Db 13529 ATGGGAGGAGAGAGAGGGGTGATTTTCCATCTGATGTGGAGTTTCAAGCTCTGTCT 13588  
 QY 1319 ----- 1319  
 Db 13589 TGAATTCTGTCTCTCAGAGGCTCAGTTTCCACTCTCAGACATGTCCAGGTGCAAG 13648  
 QY 1319 ----- 1319  
 Db 13649 AGTCTTAGTCTGACCCGACAGGCTCAAGGTGGAGTCTCTGCCCCATTCTGGGGGC 13708  
 QY 1319 ----- 1319  
 Db 13709 TCCCCAGTTCTCTCTCACTCTTCCATPACCAACAGACAAATCTAAGTGAAGCTGTCTC 13768  
 QY 1319 ----- 1319  
 Db 13769 CCATCCCCAGGGAGGAAACTGATCTCAGTCTGAGAGGAGCAAAACACCAAGTGGAA 13828  
 QY 1319 ----- 1319  
 Db 13829 GAGGAGAGAGAGGCTCCAGCCCTCCGCTGTCTGACAGCCCATCTGTGTGAC 13888  
 QY 1319 ----- 1319  
 Db 13889 ATTTAACCTTGTGGGGCCAGAGACACCCCAAGTGTCTGTGATGTGACATCGAAGGC 13948  
 QY 1319 ----- 1319  
 Db 13949 TCCCGGAGTCCGGGCTTACAAACAAGTGTCCAGCAGGCTCTGTCTTATGAAACCCAAA 14008  
 QY 1319 ----- 1319  
 Db 14009 TGGGTAGGAGAGGTGCTGGGGAATCCAGGAATTGCAAGGGAGCAGTGTGAGAGCTG 14068  
 QY 1319 ----- 1319  
 Db 14069 TGGGAGAGCGGCTCAGTGAACCAACAGGTTGCCCATGAAGGGCTGGGAGCAATAGGTC 14128  
 QY 1319 ----- 1319  
 Db 14129 TCAGAGAAAGATGCAACAGAGAGAAAGTGGGATGCTGGAGGTTTGGGAGATTAACG 14188  
 QY 1319 ----- 1319  
 Db 14189 AGGTCACGGAGGCATGGAGCTATCATGTGGTTGTGGGGGTCTCAAGTTCCCTCTCT 14248  
 QY 1320 ----- 1325  
 -----GlyGluHisCysGluVal1  
 |||||

Db 14249 GTCTCAGAGTACTTGTACTCAGCCGCTTTCCTTGTCTCAGGTTAGACACTGTAGGTG 14308  
 QY 1326 Ser1aArgSerGlyIAspCysThrProGlyValCysIAspAsnGlyIAspValAsn 1345  
 Db 14309 AGTGTCTGCTAGGCGCTGTGACCCCGGGTGTCTGCAAGAAATGGAGGACCTGTGTCAAC 14368  
 QY 1346 LeuLeuValGlyIAspCysAspCysProSerGlyIAspPheGluValProTyrCys 1365  
 Db 14369 CTGCTGGTGGCGGTTTCAAGTGCATTTGCCCATCTGAGAGCTTCAGAAAGCCCTACTGC 14428  
 QY 1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGln 1385  
 Db 14429 CAGGTGACACGCGCAGCTTCCCGCCACTCTTCATCACTTTCGCGGCTGGCCAG 14488  
 QY 1386 ArgPheHisPheThrLeuAlaLeuSer----- 1394  
 Db 14489 GCTTTCACTTCACCTTGGCCCTCTTC-GTGAAGTGTGGGCACTGGGGGTGGGAGTGGG 14547  
 QY 1394 ----- 1394  
 Db 14548 CTTGTGGGATCTGAGAGTGTCTCTGTCTGTGTGAAGTAAAGGTACAGGAGTGA 14607  
 QY 1394 ----- 1394  
 Db 14608 AGCGTGAAGACGCTGCGGACAGACCAAGTCTGCTTGAAGTGTGCTGTGTCTGCC 14667  
 QY 1394 ----- 1394  
 Db 14668 AGCAGATCTCCCTCCCTGGCTGTGTTAGGTGGCGGGGTGTGGCATCCAGGTGGTAC 14727  
 QY 1395 -----PheAlaThrIAspGluArgAspGly 1402  
 Db 14728 CCCATCTCCGCCCCCATCCCACTCTGTTCAGGTTTGCAACAAAGAGCGGAGCGG 14787  
 QY 1403 LeuLeuLeuTyrAsnGlyIAspPheAsnGlyIAspPheValAlaLeuGluValIle 1422  
 Db 14788 TTGCTGTGTACAAATGGGCGTTCCTCAATGAGAGCATCTTGTGGCCCTCGAGTATTC 14847  
 QY 1423 GlnGluGlnValGlnLeuThrPheSerAla----- 1432  
 Db 14848 CAGAGAGAGTCCAGCTCACTTCTCTGC-AGGTGATCAAGTGTCCCTCCATCTTGGC 14906  
 QY 1432 ----- 1432  
 Db 14907 CATCTTCAAAAGGCCAGTCTTCCAGCCCTGACCCCAAGCACTATCTATCAGCC 14966  
 QY 1432 ----- 1432  
 Db 14967 AANTTGGGCCCAAGCCCAAGCACTGGCACCCCAACCTGCCATTCATGACCTGTGCA 15026  
 QY 1433 -----GlyIAspThrThrValSerProPhe 1442  
 Db 15027 GGCATTCAGACTCAGCTGTCTCTTCCCGAGGAGATCAACACAGGTGTCTCCCTATTC 15086  
 QY 1443 ValProGlyIAspValSerAspGlyIAspThrValGlnLeuValTyrTyrAsnIAsp 1462  
 Db 15087 GTGCCGAGAGATCAGTATGCGCAGTGCATACGGTGCAGCTGAAATCTCAATATAG 15146  
 QY 1462 ----- 1462  
 Db 15147 GTGGGTGTGAGGACACAGAGGTTGGGGTCTGTCTTCTCAGGTGCTTACCAGC 15206  
 QY 1462 ----- 1462  
 Db 15207 CTTGAGTGCATTGCTCCAGGCTTGGGTGGCTGTGTCAGGGCATTTCTGGCTGAGAGGA 15266  
 QY 1462 ----- 1462  
 Db 15267 AGGAGCGCTGGAAAGTGCACCTTGTGGCAATTCACCCCACTGGGCACTGCTGC 15326  
 QY 1462 ----- 1462  
 Db 15327 ATGCTCTGTCTCACTAGGACCGCAACTTGGACACTGAGAAAGTTAGTATCATGCA 15386



[illegible]

QY	1595	laPheserCyrgsluCyeprolenglyPhehlylysserCyvalagln-----	1611
Db	16467	CGTCAAGCTCGAGGTGCCCTCCCTGAGCTTTGGGGGCAAGACTGCGCCAGAGTACAGAGG	16522
QY	1611	-----	1611
Db	16527	GGGCTGTTAGAGGCCACAGCTGGGGTGCATCAACAGTCTGGGAACCTGGCAGGGTTGGG	16588
QY	1611	-----	1611
Db	16587	GCAAGCAGTGGGCAAGGCTCTGCTGAGCGGGGCTGTGGGTGAAAAAGGTCTGGGCAAG	16644
QY	1611	-----	1611
Db	16647	AGCCCTGAAGAGGCAACACAGAGACAAAGGGGCACTCAGAGAGCAGCCTGGGCAAGC	16700
QY	1611	-----	1611
Db	16707	GTGGGGCGGTCCAGGCGCAAGTACGCACCTTTGAGAGGGCGGGCTGTATGAGGGAGTGGGC	16766
QY	1611	-----	1611
Db	16767	TCGTGCTCCGCTGACGCCGCCACCGCTGAGCATCAGCCCCAGGGGCTCTGAGGCTCC	16822
QY	1612	-----	1612
Db	16827	ACCGTCACAGTCTGCTTTTCGCTCCGCCCAAGAAATGGCCATCCACAGCACTTCTG	16888
QY	1621	GlyserSerLeuValAlaTTPHlsGlyLeuSerLeuProIleserGlnProTPryLeu	1640
Db	16887	GGCAGCAGCTGTGGGTGGCTGGCAATGGCTTCCTGCTGCCATTCCTCCAACTCTGTATCTC	16944
QY	1641	SerLeuMetPheArgThrArgGlnAlaAPGlyValLeuLeuGlnAlaIleThrArgGly	1660
Db	16947	AGCTTCATGTTTCGGCAAGCCCGCAGCGGCGAGGTGTCTGTGTGACGGCCATCCAGGGGG	1700
QY	1661	ArgSerThrIleThrLeu-----	1666
Db	17007	CGCAGCACATCAACCTTCACAGGTGATCATGAAAGGCGGCTGGCCCTGGCCTGGCCATA	17066
QY	1666	-----	1666
Db	17067	GGGCGCTGGTAGCTCTAGAGCGGCTGACAGAAATGGCTGGCAGGTCTGGGCAAGGGT	1712
QY	1666	-----	1666
Db	17127	GGGACATATTAGAGGCGCTGATCCGTTGGAAAGTCAATGCTGCCACTGTGGGCT	1718
QY	1666	-----	1666
Db	17187	GAGGAAATTAATACGCTTCTCTCTTATAGGCTCAGCTGGGCAAGCGGAATGGATTGAT	1724
QY	1666	-----	1666
Db	17247	GGTCACAGAAAGGGAACGTGAGCTGGAGAAACGTATTAAGGCCGTCTGCTACCTGT	1730
QY	1667	-----	1667
Db	17307	TTTCTTCTCTCCGTGGTGTACGTACGAGAGGGCCACGTATGCTGAGCCGTGGAGGGCA	1736
QY	1661	GlyLeuGlnAlaSerSerLeuArgLeuGlyProGlyArgAlaAsnAPGlyAspTTPHls	1700
Db	17367	GGGCTTACGGCTCTCTCTCTCCGTCTGGAGCCGGGCGCAATGACGTGACTGGCAC	1742
QY	1701	HisAlaGlnLeuAlaLeuGlyAlaSerGlyLeuProGlyHisAlaAlaIleLeuSerPheAsp	1720
Db	17427	CATCACAAGCTGGGCACTGGGAGGCCAGGGGGGGCCGGCCATGCTATGCTCTTGAT	1748
QY	1721	TyTIGlnGlnAlaArgAlaGlyLeuAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer	1740
Db	17487	TATGGGCAAGCAAGAGGCAAGGGCAACTGGGCCCCCGGCTGCATGGTGTGCACCTAGC	1754
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly	1760

Db	17547	AAACATACAGTGGCGGAATACCTGAGCCAGCCGCGGTGTGAGCCGTGGCTTTGGGGGC	17606
Qy	1761	Cys-----	1761
Db	17607	TGTTTGCAGTGAAGTGTCTGCTCCCTGCTCCATCCCTCCCCACACACTCGAGCCCC	17666
Qy	1762	-----Leu-GlnGlyValAlaArgValSerAspThrPr	1771
Db	17667	GCTCACTGAGGGCAACCTGCTCTGCTCCCTGAGGGTGTGGGGTGAAGCCGATAGGCC	17726
Qy	1771	OGIuGlyValAsnSerLeuAspProSerHisGlySerIleAsnValGlnGlnGlyCys	1791
Db	17727	GGAAGGGGTTTAAACGCTCGATGCCAGGCCATGGGGAAGACATCACTGAGCAAGCTG	17786
Qy	1791	ASerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAs	1811
Db	17787	TAGCTGTCTACACCTGTGTGACTCAACCCGTGTCTGTCTAAAGCTATTGCAAGCAACGA	17846
Qy	1811	PIrAspSerTyrSerCysSerCysAsp-----	1820
Db	17847	CTGGGACAGCTATTCTCTGAGCTGTGA-TCCAGGTATGCTAAGATCCAGGGCAACGGGC	17905
Qy	1820	-----	1820
Db	17906	AGGTTATCAGGTGCTGGGGCCACATGCTGGCTGTGTGTGACTGGGGCATGGGGCA	17965
Qy	1820	-----	1820
Db	17966	TCACACACCTCTGAGTCTTAAGTTCACACACAAAGCCATGACGCTGATAGGA	18025
Qy	1820	-----	1820
Db	18026	TGCTATGTGAAAGAAACAGATGAAGGTTTGATGAAGATGTGCCAGGGAGGGTGGG	18085
Qy	1821	-----ProGlyTyrTyrGlyValAsnAsnCysThrAsnVal	1831
Db	18086	GGAATGAGCCTCTGTGCTCTGTGCTGCCAGGTTACTATGTGTACAACTGTACTAATGT	18145
Qy	1831	LCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLysProSerAlaProHis	1851
Db	18146	GTCGACCTGAACCCGCTGTGAGCAACAGCTGTGTGTATCCGGCAAGCCCAAGTCCCCCA	18205
Qy	1851	SGlyTyrThrCysGlnCysProProAsnTyrLeuGlyProTyrCysGlnThr-----	1868
Db	18206	TGGCTATCTCGCAGAGTGTCCCCCAATTAACCTTGGGCATATCTGAGACCAAGGTAAAC	18265
Qy	1868	-----	1868
Db	18266	AGACACAGGCAATGTGGCAGCAGGTGCCAGTGGCTGTCTCTGTGTGTCTCTCA	18325
Qy	1868	-----	1868
Db	18326	GAGCCCGAAGCGTGGCTATCCACAGCCAGGGTCAGAAAGGSCACATAGGGCTCACT	18385
Qy	1869	-----ArgIleAla	1871
Db	18386	AGGTTAGTGAAGTGGCAGGGGAGCTCATGCTTACCTGGGTCCCTCTGCAAGGATTG	18445
Qy	1871	SPGlnProCysProArgGlyTyrPrTPGlyHisProThrCysGlyProCysAsnCysAspV	1891
Db	18446	ACCAAGCTTGTCCCGTGGCTGTGGGAGCATCCCAATGTGGGCCATGACATCTGTATG	18505
Qy	1891	AlSerLysGlyPheAspProAspCysAsnLysThrSerGlyGlyCysHisCysLys-----	1909
Db	18506	TCAGCAAAAGGCTTTGACCCACAGCTGCAACAAGACAGCGCGAGTGCACACTGCMAAGTGA	18565
Qy	1909	-----	1909
Db	18566	CAGCCCCAAGCAAGCTTCACATGTGGCACTTGGGCTCTGTCCACATCTCTGGGCCCT	18625
Qy	1909	-----	1909
Db	18626	AGAGGGCAGGCTGTATGAGGATGGGCTTCTCTGTATAGCTGAGGGCCTTGTGCTTGTCT	18685
Qy	1909	-----	1909
Db	18686	CACAGCAGGGCAGTGTGCTGCTTACCTTTCGTGTGTCTGAGGCGCAGGGGACCTTGAGC	18745
Qy	1909	-----	1909
Db	18746	CAATGTATACCAATGAGCTCTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	18805
Qy	1909	-----	1909
Db	18806	TGTGACATCTCGCTATATGGGCTCAACTCCAAAGCATTTGCTATGTATATTATTTGAATA	18865
Qy	1909	-----	1909
Db	18866	CTGCTCTGAGGGGACCTGAGGTGGGACCTCTCCACAGGGGCCACAGCTGAGGAAGAC	18925
Qy	1910	-----G	1910
Db	18926	CCATGACATCTGGGAAAGAGCCTGGCCAGTACACCGTTTCTTCCCTTCTCTCAGG	18985
Qy	1910	IuAsnHisTyrArgProProGlySerProThrCysLeuLeuCysAspCysTyrProThrG	1930
Db	18986	AGAACCACTAACCGGCCCCCAAGGACCCCACTGCTCTTGTGTACTATGCTACCCCAAG	19045
Qy	1930	IySerLeuSerArgValCysAspProGlyIuAspGlyGlyCysProCysLysProGlyValI	1950
Db	19046	GCTCTTGTCCAGAGCTGTGTGACCTGTGAGAGTGCAGAGTCCAGTCAAGCCAGGTTCA	19105
Qy	1950	IeGlyArgGlnCysAspArgCysAspAsnProPheAlaGlyValThrThrAsnGlyCysG	1970
Db	19106	TGGGGGTGAGTGTGACCGCTGTGACAAACCTTTGTGTAGGTCAACCAATGTCTGTG	19165
Qy	1970	Iu-----	1970
Db	19166	AAAGTGGGCTCTGGGATGGGTGGCAGCCTCTTCAAGTGTCTAGGACCTGACCC	19225
Qy	1970	-----	1970
Db	19226	CCAGAAAGCCAGAAAGGGCTGTGTGACAGGCTGTGGGAGGCTTGGGAGGCTTAGGAG	19285
Qy	1970	-----	1970
Db	19286	GGCTGGGAGCTGGGCAAGACAGAGAGGAGAGCTCTCTGTGTGACCATGTGCTCT	19345
Qy	1971	-----ValAsnHisTyrAspSerCysProArgAlaIleGlyIuAlaGlyIleTyrTPProA	1988
Db	19346	TCCCGCACTGATATATGACAGCTGCCACAGCAGCATTTGAGGCTGTGGATCTGTGGGCC	19405
Qy	1988	rgThrArgPheGlyLeuProAlaAlaAlaProCysProLysGlySerPhe-----	2004
Db	19406	GTAACCGCTTGGGGCTGCTGTCTGTCTCTCTGTCCCAAAAGGCTCTTGTGTAGTGT	19465
Qy	2004	-----	2004
Db	19466	GGAGGCCGATGTGATGTGAGACATGAGCTCTGTGTATGTGGATGAGGAGGAGA	19525
Qy	2004	-----	2004
Db	19526	GCTCTGGGTGAAGATCCGGGGACCCCACTTCTGCGCTCCAGATCTGGGATTTCAATT	19585
Qy	2004	-----	2004
Db	19586	CCCATCTTCCTTGTATCAATTCCTCCCTTCCCATGTGGCCCAAGGTGTGCTTACATCTTG	19645
Qy	2004	-----	2004
Db	19646	GACTGTTTGTGTTTGTGTTGTTGTTTCTTGTGGAGCAGAGTCTGTACTGTACAC	19705
Qy	2004	-----	2004
Db	19706	CAGGCTGTGTGACAGTGGCGTGTACTCGCTGCTGACCACTTGCCTCCAGGTTCAAG	19765

QY 2004 ----- 2004  
Db 19766 CAATTTTCCTGCTAAGCCTCTGCGCCGACCAACCTGGCTAATTTTTTTTTTTTTT 19825  
QY 2004 ----- 2004  
Db 19826 TTTTTTTTGTATTTTTAGTAGAGACAGGGTTTACCATGTTGGCCAGTCTGGTCTCGAA 19885  
QY 2004 ----- 2004  
Db 19886 CTCCTGACCTCATGATCCGCCCTCGCCCTCCCAAAGTCGGGATTTACAGGGGTGAG 19945  
QY 2004 ----- 2004  
Db 19946 CCACTGACCCCAACCAATTTTGAATTTTTTTTTTTTTTTTGAAGACGAGTCTGCTC 20005  
QY 2004 ----- 2004  
Db 20006 TGTCAATCAGGCTAACATGACATGGCATGATCTCGGCTCTGCGAACCTCCACCTCACAG 20065  
QY 2004 ----- 2004  
Db 20066 TTCAAGCATTTCTCTACTGACCTTCCGAGTACTGTACTACAGGGGTGTGCTACA 20125  
QY 2004 ----- 2004  
Db 20126 TGCCAGCTAATTTTTTATATTTTTTTAGTAGAGACAGGGTTTACCATGTTGCCAGGA 20185  
QY 2004 ----- 2004  
Db 20186 TGCTTCGATCTCTGACCTTGATGATCAGCCGCTCGGCTCCCAAAGTGCAGATTA 20245  
QY 2004 ----- 2004  
Db 20246 CAGGGGTGAGCAACGCTGCCAGCCCAATCTTGAATCTTAACAAAAATGCTGAAAAAC 20305  
QY 2004 ----- 2004  
Db 20306 CTGCACTGTCTTGGCATAGGGCTGTGTAGAGATTCAAGAATGATCATGAGGGTTCC 20365  
QY 2004 ----- 2004  
Db 20366 AATCGTCACTCAACCAAGTCTGAGTTAGCCACCACAAAGGACAGACGTCTGGGG 20425  
QY 2004 ----- 2004  
Db 20426 AGTAGGGGGTGTGGCTGTGAACCCAGAACCCAGCCAGAGTGGGGCTGGCGCTGGT 20485  
QY 2004 ----- 2004  
Db 20486 CCTGGACTGGCAATGAGCAGGTTTATAGGACAGAGTAGCGGAGACAAAGTGGTGC 20545  
QY 2004 ----- 2004  
Db 20546 TGGAGACAGGACAGGATGGGGGCTCAAGAGAGTGTATGCCCTGTTCCCTTAAGCCCATG 20605  
QY 2004 ----- 2004  
Db 20606 TGTGACATGATCTTGGGAGATTTTTTGGCCCTGGTGTCTACATACCTGTCTTCCGTTT 20665  
QY 2004 ----- 2004  
Db 20666 CACCAAGGCGCTTGTGTGCTAGTACTGTCCCGGCCATGAGGTGTGGCTCCAGCAG 20725  
QY 2004 ----- 2004  
Db 20726 AACCTCTCAGCCAGTCTCAGGGGCTTCTGTGTGTCTCCCTTAGAGCCGCGCTTTTGGCT 20785  
QY 2005 ----- 2021  
Db 20786 TCCTTCCCCCAGGAGATGCTGTGCGCCACTGTGATGAGCACAGGGGGTGGCTCCCCCAA 20845

QY 2021 snLeuPheAsnCyethrSerIleThrPheSerGluLeuLysGlyPhe----- 2036  
Db 20846 ACCCTTCAACTGACCGTCCATCACTTCTCAGAACTGAAGGGCTTCGTAAATGAACCCC 20905  
QY 2036 ----- 2036  
Db 20906 CTCACTTCATCTTTTCCCTGTCTCTGCTGAGTCTCACTTGCCCCCACTCCCATCTT 20965  
QY 2036 ----- 2036  
Db 20966 TGAGAACGGGGCTTCTGGAATTCAGCCTGTGTCTTCTGGGGCTCCACGCTGAGAGGCC 21025  
QY 2036 ----- 2036  
Db 21026 GTCTCTCACTCTGAGGCTCTTGTGCTCAGAGTCTCGCTCACTCTGCTCTCCG 21085  
QY 2037 ----- 2051  
Db 21086 CTCGGTCTGTCTGAGGCTGAGCGGCTACAGCGGAATGATGAGGCTTAGACTCAGGGCGC 21145  
QY 2052 SerGlnLeuAlaLeuLeuLysAsnAlaThrGlnIleuAlaHisGlnSerThrGln 2071  
Db 21146 TCCAGAGAGTACGCTGCTCTGCGGACGCGACAGACACAGCTGGCTACTTGGGC 21205  
QY 2072 SerAspValIleValAlaIleThrGlnLeuAlaThrArgLeuLeuAlaHisGlnSerThrGln 2091  
Db 21206 AGCAGCTCAAGTGGGCTTACCACTGCGCCACGCGGCTGCGGCCACAGAGACCCAG 21265  
QY 2092 ArgGlyPheGlyLeuSerAlaIleThrGlnAspValHisPheThrGlu----- 2106  
Db 21266 CGGGCTTGGGCTGTCTGACACAGAGCTGTCACTGAGGTGGGCTTGAGGA 21325  
QY 2106 ----- 2106  
Db 21326 TGCAAGGCTGCTGTGTAGATAGGGGTCAATGTAGTGAAGCTGTGTATGCAACTGG 21385  
QY 2106 ----- 2106  
Db 21386 GGGCAGAGGGGCGCTCCATCCACTCAAGAGAGCTCCCTGTGGAGAGGCTTCA 21445  
QY 2106 ----- 2106  
Db 21446 TGAACCTGTGACCCCTGCGCCAGCCCTTCCCAACCTTTCATATGTAATATGTG 21505  
QY 2106 ----- 2106  
Db 21506 TGTGGGCGATGTCTGACCCAGACAGAGGCTGCTGTGTACATGGCTGGAGGA 21565  
QY 2106 ----- 2106  
Db 21566 AGCAGTATCTGAGCATGTGTGTGGGTCCAGGACAGGGCTGGAGACTTATATAGA 21625  
QY 2106 ----- 2106  
Db 21626 GAATGAGAGAGGGCTTAAGGGGAGAGCATCCAGTGACTGATGTGGCAATGGCTTCA 21685  
QY 2106 ----- 2106  
Db 21686 GGAGCATTTCAAGGAGGCAACAGTGAACACAGAGGAGGGCTGTGGGAGAGGGA 21745  
QY 2106 ----- 2106  
Db 21746 CTGGCCGGGACAGGCGGGGCTTCATGTGGGAGTCAAGTGTGGTGTGACACA 21805  
QY 2107 ----- 2118  
Db 21806 CCAACCTGACCTTGGCCACCCCAAGAACTGTCTGGGGTGGGAGCGCCCTCTGAGACAC 21865  
QY 2118 rAlaAsnLysArgHisThrGluLeuIleGlnIleThrGlnIleGlyIleThrAlaIlePleuLe 2138  
Db 21866 AGCCCAACAGGCGGCACTGGGAGCTGATCCAGACAGAGGGGTGGCACCGGCTGGCTCT 21925  
QY 2138 uGlnHisIleThrGluAlaIleAlaSerAlaLeuAlaGlnAsnMetArgHisThrIleLeu 2158

```

Db 21926 CCAAGCATATGAGGCTTACGCGCAAGTCCCTGGCCCAAGACATGCGGCACACTTACTAAG 21985
Qy 2158 rProphetHrilleValThrProhmiLe----- 2167
Db 21986 CCCCTTACCAATCTGTACCCCAACATTTGTAAAGCTGTGCTGGGTTGGGAGGGGTT 22045
Qy 2167 ----- 2167
Db 22046 TGTGAGGAGATCCCCGACAGAGCGGCTGTGGGATCTGCTGCTCAAGGTCCG 22105
Qy 2168 -----ValIleSerValValArgLeuAspLysGlyAsnPha 2180
Db 22106 ATCTGTACCATCCCTTCTTATGATCTTCGTAGTGTGGCTTGACAAAGGAACTTTG 22165
Qy 2180 IagIValAlaYleuProArgTyrGluAlaLeuArgGlyGluGlnProProAspLeuGluT 2200
Db 22166 CTGGGGCCAAAGCTGCCCGGCTTACAGAGCCCTGCGTGGGAGACGCCCGGACCTTGAGA 22225
Qy 2200 hThrValIleLeuProGluSerValPheArg----- 2210
Db 22226 CAACAGTCATTCCTGCTGAGTGTCTTTCAG-AGGTCAAGTGTGGCCATGATGAGTTG 22284
Qy 2210 ----- 2210
Db 22285 GGAAGTGAACCCAGTGTCTGTGCAGACTCCACAGAGCAGGGCCCAAGCTAAGTGAAC 22344
Qy 2211 -----GluThrProProValValArgProAlaGlyProGlyGluAlaGln 2226
Db 22345 AGTGTCCCTCCCAAGAGAGCCCGGCTGTGTCAAGCCCGCAGGCCCGGAGAGGCCAAG 22404
Qy 2226 IuProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGlyAlaVal 2246
Db 22405 AGCCGAGAGAGCTGGCAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 22464
Qy 2246 aAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAlaThrTrpPro 2266
Db 22445 TGGCAGAGGTATCATCTACACGCACTGGCCGGCTACTGCTCATTAATGACCTTG 22524
Qy 2266 sPlyArgSerLeu----- 2270
Db 22525 ACAAGCGAGGCTTGAAGGTACAGAGGAGGAGCAGGTGTGGGTAGGGGTATGGGTGGGC 22584
Qy 2270 ----- 2270
Db 22585 GGTGAGTCTGAGGATGAGAGGGGAGTGGGGGCGTCTCCCAAGTCATGACTGCGGTG 22644
Qy 2271 -----ArgValProLysArgProIleIleAsnThrPro 2281
Db 22645 TGACTGTGCACTGACTGCGCCCAATCAAGAGTCCCAACCGCCGATCATCAACACACC 22704
Qy 2282 ValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysPro 2301
Db 22705 GTGGTGAACATCAAGGCTCATATGATGAGAGCTTTCGCCCGGGCCCTTGACAAACCC 22764
Qy 2302 ValThrValGlnPheArgLeuLeuGluGlnThrGluGlnArgThrLysProIleCysValPhe 2321
Db 22765 GTCAAGGTGCAAGTTCGCCCTGTGAGAGACAGAGAGCGGACCAAGCCCATGTGTCTTC 22824
Qy 2322 TrpAsnHisSerIleLeu----- 2327
Db 22825 TGAACCATTCATCAATCTGTAGAGCTGCACTGCGCCCTCAGGCTTCGGGCTGAA 22884
Qy 2327 ----- 2327
Db 22885 GTCCAGGCCCCGATGCTCAACCTCGTGTCTCCCTGACCCCTGCTTCTCAACAGG 22944
Qy 2338 ValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAspGluSer 2347
Db 22945 GTCAAGTGCACAGGCTGTGGCTGGCCAGAGGCTGTGAAGTGTCTTCGCAATGAGAGGC 23004
Qy 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValIleMetAspValSerArg 2367

```

---

```

Db 23005 CAGCTCAGCTGCAGTGCACACCATGACAGAGTTGCTGTGTCTATGAGACGTTTCTGG 23064
Qy 2368 ArgGlu----- 2369
Db 23065 CGGGAGGTGGGGCCCAAGGGGCAAGTGCAGAGCCGTGGTGGGCACCCAAGGGCAAGGG 23124
Qy 2370 -----As 2370
Db 23125 CTGGGTGCTCAGGTCTGCTCCCTTCTATTCCTCTGGCCCCCTGCCACCTTACTTGACAA 23184
Qy 2370 nglyGluIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaI 2390
Db 23185 TGGGAGATCTCTGCACTGAAGACACTGACATAGGTGGCTTAAAGTGCACCTTGGCTGC 23244
Qy 2390 aLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGly 2410
Db 23245 CCTTGTGCTCAGCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 23304
Qy 2410 YIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyI 2430
Db 23305 CATCCGAGCTAACCTGACAGCTGCCCTGGGCTCAGCTGTGCTTCTCTCTGGGAT 23364
Qy 2430 eAsnGlnAlaAspLeuPro----- 2436
Db 23365 CAACCAAGCTGACCTCTCTGTAGATGCTCTTACTGCCCAGAACTGTCCCCACCTTCTC 23424
Qy 2436 ----- 2436
Db 23425 AGCGCGCTCCCAAGCCCCCACTGGCAACCTGTCTCTGCAACATGAACTGTATTAAG 23484
Qy 2436 ----- 2436
Db 23485 TGCCTAGTGCAGCACTGGGCCAGAGGTTTCTCTTGTGTGCTCCCCGGGATCCCCAG 23544
Qy 2436 ----- 2436
Db 2436 ----- 2436
Qy 2436 ----- 2436
Db 23545 CACCTGCTTGCCAGGCTTCCCTGGAAGACAGTCCCAACCCAGGCCCTCTCCAT 23604
Qy 2436 ----- 2436
Db 23605 GCCTGACCCCAAGCAGAGCTGTGCTGGGCGGGCCCCGGTCCGTGACCTGCGCC 23664
Qy 2437 -----PheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysT 2453
Db 23665 TGGCCCTCAGTTTGGCTGCACAGTCAATGCAATCTGTGCACTTCTGTACTGCA 23724
Qy 2453 hrPheSerTrpAlaLeuLeuGluAlaLeuHisIleLeuTyrArgAlaLeuThrGluValArg 2473
Db 23725 CTTTCTCTGGGCTCTGTGTGAGGCTTGCACCTGTACCGGGCACTCATAGAGGTGGCG 23784
Qy 2473 sPValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheI 2493
Db 23785 ATGTCACACCGGCCCAATGCGCTTCTACTACATGTGGGCTGGGGGCTGCTCTTCA 23844
Qy 2493 IeThr----- 2494
Db 23845 TCACAGTACTCCACCATTCACAGTCTGGGGTCCACATCCGTGGTCCACTTTGT 23904
Qy 2494 ----- 2494
Db 23905 GCCATGTTCTCCACCAATACAGAGCCCTGAGGCCCAATCCCATGCCCCAGGCG 23964
Qy 2494 ----- 2494
Db 23965 CTTATTCAGAGGTGTCCTGTGTTTAACCAAGCTGTGACGCCGACCCAGGCGTCA 24024
Qy 2494 ----- 2494
Db 24025 CTCCCTTATCTGGAGATCGGTAGGGGCGCATGTGTGGCAGAACCTTTTCCATGCTT 24084
Qy 2494 ----- 2494
Db 24085 CATGCTGGCCCTGTGAGCCCACTGCGCGAGCCCTTATTCAGAGCCCTCTATCAACC 24144

```

QY 2495 -----GlyLeuAlaValGlyLeuAspProGlnG 2504  
 Db 24145 CTCACGTGCTCCGCTGCTCTCCATGCTCCAGGGCTA-GCCGTGGGCTGGAGACCCGAGG 24203  
 QY 2504 LYTGTGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIleTrpSerPheA 2524  
 Db 24204 GCTACGGGAACCTTGACTTCTGCTGGCTCCATATGACAGCTCATCTGAGATTG 24263  
 QY 2524 LAGlyProValAlaPheAlaValSer----- 2532  
 Db 24264 CTGGCCGGAGGCTTGTGCGCTCTCGGTAGTGTAGCAGGTGGGTGGGTGTCACCTGT 24323  
 QY 2532 ----- 2532  
 Db 24324 GGCCCTCTTGTGCTGCTCTCTGCTGCGGGCGGGTGTCTCAGAGACTCCCTTATAGA 24383  
 QY 2532 ----- 2532  
 Db 24384 ACAGCTGAGGCCACGGGGCCCTGTGGGCTGGGTGGAAGCTGTTGTCCCACTGACACC 24443  
 QY 2533 -----MetSerValPheLeuTyrIleLeuAlaAlaArgAlaS 2545  
 Db 24444 CCATGCCCTTGCCTCCCTGCTAGTAGTGTCTTCTGTACATCTGGCGGCCCGGGCT 24503  
 QY 2545 erCysAlaAlaGlnArgGlnGlyPheGlnIuLySerGly----- 2557  
 Db 24504 CCGTCTGCTGCCACGGCGAGGGCTTTGAGAAAGAAAGTCTGTGTAGATATAGGTTGGG 24563  
 QY 2557 ----- 2557  
 Db 24564 GTGCCGTGGGCATGGGACGACACAGCATGGAGGTCTCATGGCCGAGCTCATGGCCTGT 24623  
 QY 2558 -----ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaT 2575  
 Db 24624 CCTATCCCTCAGC-TGGGGCTTGAGCCCTCTTCCGCGCTCTCTGCTGCTGAGGCA 24682  
 QY 2575 hrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheIleTyrLeuPheA 2595  
 Db 24683 CGTGGCTGGCTGCACTGCTCTCTGTCAACAGCAGACCCCTCTTCACTACTCTTTG 24742  
 QY 2595 LahrCyAsnCySile----- 2600  
 Db 24743 CTACCTGCAATTCATTCACAGTACCTGGCCAGCCTGTGGAGAGGACCTGGGCT 24802  
 QY 2601 -----GlnGlyProPheIleP 2606  
 Db 24803 GTGGATGCCGATATGACACAGACCGTGTGCTCTTGTGCTGCCAGGGCCCTTATCT 24862  
 QY 2606 heLeuSerTyrValIleLeuSerIleGlnValArgIyAlaAlaLeuLySLeuAlaCysSerA 2626  
 Db 24863 TCCTCTCTTATGTGTGTCTTACAGACAGGAGGTCCGGAAGACACTCAAGCTTGGCTGAGCC 24922  
 QY 2626 rglYsProSerProAspProAlaLeuThrIlySerThrLeuTrpSer----- 2642  
 Db 24923 GCAAGCCCAAGCCCTGAGCCCTGTCTGACCAAGTCCACCTTGACCTCGGTAGGAGAC 24982  
 QY 2642 ----- 2642  
 Db 24983 CAGGGGTCTCAGAGAGGCGGTGAAGGAGGAGGAGGAGGCGCAAGTAGCCTTACT 25042  
 QY 2642 ----- 2642  
 Db 25043 TTTGGCCCACTCCCTTTCTCTTCTTCATCCCTTCTGTGAAGGTGGAAGGAGAGGTG 25102  
 QY 2642 ----- 2642  
 Db 25103 AAATGCTGTGTGGGCTGGGAGAGAGATGTGGAAGGGTGTGTAGATTTTGACTTGA 25162  
 QY 2642 ----- 2642  
 Db 25163 GGAGATGGGGCTGAGACAGATATATCAGTATATAGTCTTGCACCTGAGATCTCTCA 25222

QY 2642 ----- 2642  
 Db 25223 TGAATTTTAGTAATTTTCTGTCTCCAGAGGTCTCTGAGAGACAATTTCTGTCTCAT 25282  
 QY 2642 ----- 2642  
 Db 25283 CTTTATACAGACCATATATATAGTAGTCAGAGATCCCTGGGGGATGACAGGGGTAGACA 25342  
 QY 2642 ----- 2642  
 Db 25343 GCTGTGGCAGAGCTGTGACCACTGCCAGGCTCAATATGGGACAGCGGAGAGAGGAG 25402  
 QY 2642 ----- 2642  
 Db 25403 CCTCTTTACAGACTGTAGCATGTGGAAGGTGAACTTAGGGCAAGTTCCCTCCACCTC 25462  
 QY 2642 ----- 2642  
 Db 25463 CTTCTTCCAGCCTGGGGCAGCCATCCCACTCCCACTTACTGACTCTCTGTCCCTG 25522  
 QY 2643 -----SerTyrAsnCyAspProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGlyA 2661  
 Db 25523 CCTAGTCTTACACTGCCCCAGCCCTTACGAGATGGCGGCTGTACAGCCCTTACGAGG 25582  
 QY 2661 psSerAlaGlySerLeuHisSerThrSerArgSerGlyLySerGlnProSerTyrIleP 2681  
 Db 25583 ACTGGCGGGCTCTCTGACACAGCACAGTGTGCTGGGCAAGAGTCAAGCCCACTATCC 25642  
 QY 2681 rOPheLeuLeu----- 2684  
 Db 25643 CTTCTTCTGAGTGAATCCCGAGATGGAGGTGAGAGGAGGAGGAGGCCACG 25702  
 QY 2684 ----- 2684  
 Db 25703 CATGTGACCCAGGCCAGCCAGCTGTGTGGAGTTGAGAGACACACTGTGTGTAAGT 25762  
 QY 2684 ----- 2684  
 Db 25763 GGGGGCAGCTTGATTAAGAGTGTAAAGACCCACAGCAGAAACAGATCCCAAGGG 25822  
 QY 2684 ----- 2684  
 Db 25823 AGAGAGAGACTGGGACCTTGGGCAAGGGGCCAGGCTGACCCCTTCAGATGTCTCATC 25882  
 QY 2685 -----ArgGlnGlySerAlaLeuAsnProGlyGlnGlyProProGlyLeuGlyAspProG 2703  
 Db 25883 TTCATAGGAGAGAGTCCGCACTGAACCTTGGCAAGGGCCCTTGGCTGGGGATCCAG 25942  
 QY 2703 lySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAsp----- 2715  
 Db 25943 GCAAGCTGTCTGTGAAGTCAAGACAGCAGCATGG-TGAGACAGAACGCTTGAGCA 26001  
 QY 2715 ----- 2715  
 Db 26002 CCCAGAGGGCAGTGTGGCTGTGCTTTTACTGAAGTGGGTGAGAGTGGCTGGGCTGT 26061  
 QY 2716 -----ProAspThrAspSerAspSerAspLeuSerLeuGlnA 2728  
 Db 26062 GATCTCTCCGTGGCTCTTAATCTCTGACAGGACTCCGACAGTGACTGTCTTTAGAG 26121  
 QY 2728 sPAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnG 2748  
 Db 26122 AGACCAAGGTGGCTCTTATGCTCTTACCACTCATGACAGTGAAGGAGAAAGAGG 26181  
 QY 2748 LngLngLngLngLngLngAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGlyProG 2768  
 Db 26182 AGGAGGAAGAGAGGCCCTTCTCTGAGAGAGAGGCTGGGATAGCTTGGGGCTG 26241  
 QY 2768 lyAlGlnArgLeuProLeuHisSerThrProLyS----- 2779  
 Db 26242 GAGCAGAGAGACTGCCCTCTGACAGTACTCCCAAGGTGGGCCAGCACTGGGGCTGTGGC 26301  
 QY 2779 ----- 2779

Db	26302	CTTTGGGCGCAGTGGAGAGACAGTGGGCCCTGGGGTTCTTGGAAGAAGACTGGGTGTGT	26361
Oy	2780	-----	AapGIYg 2782
Db	26362	GGCCTCTGAGCTATCTGCCCCCTGTGTGGGCTCATCTTCCTTTCCCAACGAGTGGGG	26421
Oy	2782	lyProglYProglYLyvAlAProtrPrProglYaSPheglYTHrThralAvagIusets	2802
Db	26422	GCCCAAGGCGCTGGCAAGGCCCCCTGGCCAGAGACTTTGGGACCAAGCAAAGAGAGTA	26481
Oy	2802	ergIyaengIyAlAProglIngIuaRgLeuarvgIuaengIyaApAlaleuSerArGlug	2822
Db	26482	GTGGCAAGGGGGCCCCCTGAGAGAGCGCTGGGGGAGATGAGATGCCCTGTCTCGAGAGS	26541
Oy	2822	IysertleungIyProleuProglYsefserAlaglInPohIslyS-----	2836
Db	26542	GSTCCCTTAGGCCCCCTTTCAGGGCTCTTCTGCCAGCTCCAAGGTGAGTGGGGCACCC	26601
Oy	2836	-----	2836
Db	26602	CCACTGCCCAGCTCCCCCTACTAGCAGAGCTCATCTCAACTTCCTCTGTGGCCGACCC	26661
Oy	2837	-----GlylleuIyIslyIslyCySleuProthrIIeserg	2849
Db	26662	TCAcAGCCCCGCCCCCGCCCAAGGCACTCTTAAAGAAAGTGTCTGCCACATCAAGGS	26721
Oy	2849	IuYsserSerleuLeuAArgLeuProleuIngIncYethrNgIyserseraRgIysers	2869
Db	26722	AGAAAGACAGACTCTCTCGGCTCCCCCTTGAGCAAGTGCACAGGGGTCTTCCCGGGGCTCT	26781
Oy	2869	eralaserGluIySerArGlYglYproProProArGPoPoProArGlnSerLeug	2889
Db	26782	CCGCTATGTAGAGGAGCGCGGGAGAGCCCCCTCCCGCCCAAGCCCCGGCAGAGCTTC	26841
Oy	2889	InglIngIueAenGIyValmetProIleaIameterIleylaagIyThrValaspG	2909
Db	26842	AGGAGCAGCTGAGAACGGGGTCAATGCCCATGCATGAGCATCAAGAGCAGCAGCGTGGATG	26901
Oy	2909	IuaPsserSerGIySerGIu 2915	
Db	26902	AGCACTGCTCAGGGCTCCAG 26921	
RESULT 12 AC106175/c			
LOCUS		235320 bp DNA linear HTG 13-MAY-2003	
DEFINITION	Rattus norvegicus clone CH230-7M10, WORKING DRAFT SEQUENCE, 4		
ACCESSION	AC106175		
VERSION	AC106175.5 GI:30578758		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	1. (bases 1 to 235320) Murphy,D.,Marlee,M.,Metzker,M.,Lee,S.,Abramson,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Aishbrooks,S.,Amin,A.,Anguilano,D.,AyalaDebeli,V.,Aoyagi,A.,Ayodeji,W.,Baca,B.,Baden,H.,Baldwin,D.,Bandaranaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,Biswalto,K.,Blair,J.,Blankenburg,K.,Blych,P.,Brown,M.,Bryan,N.,Buhaý,C.,Burck,P.,Burrell,K.,Calderon,B.,Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Chen,H.,Chen,Z.,Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,X.,Cleaveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Crete,A.,D'Souza,L.,David,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dedertich,D.,Delgado,O.,Demson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,Fernandez,S.,Finlay,M.,Flagg,N.,Forbes,L.,Foster,M.,Poster,P.,Fraser,C.M.,Gabriel,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,Gebreyesogie,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,		

Gharate, P., Haaland, M., Hamli, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.  
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshewa, L., Louised, H., Lozato, R.J., Lu, X., Ma, J.,  
 Meeshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhney, S., McLeod, M.P., McNelli, T.Z., Meenen, E.,  
 Mlagaivley, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgano, M., Morris, K., Morris, S., Mundaasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nockelmech, O., Okunou, G., Olarunpaseon, A., Pal, S., Parks, K.,  
 Paternak, S., Paul, H., Perez, A., Perez, B., Plankoch, C.,  
 Plopper, F., Polndexter, A., Popovic, D., Primus, B., Pu, L.,  
 Pizzo, M., Quiroz, T., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,  
 Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,  
 Sneed, A., Sodergren, B., Song, X.-Z., Sorrell, R., Sosa, J.,  
 Steinhilber, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Williams, R., Wleciyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, Q., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Neiderhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
 Weinstein, G., and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 235320)  
 Worley, K.C.  
 Direct Submission  
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 235320)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced g1:25007950.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are oriented and oriented, and separate  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: G4O  
 Center clone name: CH230-7N10  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0:  
 Consensus quality: 225681 bases at least Q40  
 Consensus quality: 227443 bases at least Q30  
 Consensus quality: 229154 bases at least Q20

Estimated insert size: 234593; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 231058: contig of 231058 bp in length  
\* 231059 231158: gap of unknown length  
\* 231159 232689: contig of 1531 bp in length  
\* 232690 232789: gap of unknown length  
\* 232790 233881: contig of 1092 bp in length  
\* 233882 233981: gap of unknown length  
\* 233982 235320: contig of 1339 bp in length.

## FEATURES

source  
1. 235320  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-7N10"  
misc\_feature  
1. 1015  
/note="wgs\_contig"

BASE COUNT 59423 a 53960 c 55203 g 61642 t 5092 others  
ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 235320  
Score: 12001.00 Matches: 2789  
Percent Similarity: 38.64% Conservative: 47  
Best Local Similarity: 38.00% Mismatches: 75  
Query Match: 77.20% Indels: 4436  
DB: 2 Gaps: 33

US-09-916-849a-3 (1-2923) x AC106175 (1-235320)

QY 1 MetArgSerProAlaThrGlyValProLeuProThProProProLeuLeuLeu 20  
DB 222169 ATCCGAGCGCGGCTGCGCAGCGCCCTCCCAACGCACTGCTGCGCTACGCTG 222110  
QY 21 LeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
DB 222109 TTCTGCTGCTGCGCGCGCTGCGCACTAGAGATCAAGTGGGCGCTGCTCTCTG 222050  
QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysProSer 60  
DB 222049 GGGTCGGGGGAGCGCGCTCTCTGCGGCGCTGCGCCCGGTGGCTGCTGCCAGCG 221990  
QY 61 SerAlaSerLeuLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyTyrGluLeuThr 80  
DB 221989 TCGGCTTCGAACTCTGCGCTCTACACCGCGCTGCAAGATTCGCGCATTTAGCTGACC 221930  
QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysArgProGlnSerGluAlaHis 100  
DB 221923 GGGCACTGCTGCGCCATTCAGATGCGCTAGAGGTTTGATGTCAGATACAGGGGCTCAT 221870  
QY 101 IleProLeuProProAlaProGlnGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
DB 221869 ATCCCTCTTCGCGCATCTCTGAAAGGCTGCGCCCTGAGGCTGTGACCTCTGGGTATCGGA 221810  
QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlnGlnHisProCysLeuValAla 140  
DB 221809 GGAACCTTTCTTCACAAAGGCAAGCTGACCTCTGCTCAAGAACACCTTGTAAAGGCC 221750  
QY 141 ProArgLeuArgCysGlnSerCysGlyLeuAlaGlnAlaProGlyLeuArgAlaGlyIle 160  
DB 221749 CCACGGCTCAGATGCCAGTCTTTCAGAGCTGCGACAGGCCCAAGGCTCAGGGCTGGGGAA 221690

QY 161 ArgSerProGlnGlnLeuLeuGlyValArgArgValAspAsnValAsnThrAlaProGln 180  
DB 221689 GGATCACAGAGAGATCTATGGGTGGCGCAGAAAAAGATGTGATACGCTCCCG 221630  
QY 181 PheGlnProProSerSerTyrGlnAlaThrValProGlnGlnProAlaGlyThrProVal 200  
DB 221629 TTCAGCCTCCCAAGCTACAGAGGCCAGATGCCCGGAACACAGCCTGTGCTCTGTT 221570  
QY 201 AlaSerLeuValAlaIleAspProAspGlnGlyValAlaGlyArgLeuGlnTyrThrMet 220  
DB 221569 GCTTCCTTAAGGCCATTGATTCGATAGAGGTGAGCGGGTCACTTGAGTACCAATG 221510  
QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
DB 221509 GATGCCCTCTTTGATAGCCCGCTCCCATATTCTCTCTTGAGCCCAATCATCTGTGT 221450  
QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyThrTyrSerThrHisValPheArgValThr 260  
DB 221449 GTCAACACAGCTGAGAGACTGATCGGAGACCAAGACACCAATGCTTCAGGGTCAAC 221390  
QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
DB 221389 GCACAGATCATGTATGCCCGCAGAGTGCCTTGCCACCTTATCTTATGTACT 221330  
QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlyTyrGlySerLeuArgGlu 300  
DB 221329 GACACCAATGATACAGACCTCTTTTATGACAGCAAGATTCAGAGAGAGCTCAGGGAG 221270  
QY 301 AsnLeuGlnValGlyTyrGlnValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
DB 221269 AACCTGAGGTGGCTATAGAGTGTCTTACAGTCAGAGGCGCACCGAGGGATGCCCTCT 221210  
QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlySerProSerGluValPhe 340  
DB 221209 AATGCCAATATCTATACCGCTGCTGAGAGGGCGCTGAGCGAGCCCTCAGAACTTT 221150  
QY 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluVal 360  
DB 221149 GAGATCGATCTCGCTCGGGGTATCCGAACCGCGCCCGGTGATCGGGAGAAAGTG 221090  
QY 361 GlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
DB 221089 GAATCTCAAAATATATATGAGGAGGAGTACCAAGGTGCGGACCCAGGCCCAAGAGC 221030  
QY 381 ThrThrAlaAlaValPheLeuSerValGlnAspAsnAspAsnAlaProGlnPheSer 400  
DB 221029 TTCAAGCTGCTGCTCTCTCTGTGAAATGATATGACAGCCGCCCAAGTTTAC 220970  
QY 401 GlnArgArgTyrValGlnValArgGlnAspValThrProGlyAlaProValLeuArg 420  
DB 220969 GAGAAGCGCTACGTGTTCAAGTGGGAGAGATGACCCCAAGAGCCCGTCTCCGA 220910  
QY 421 ValThrAlaSerAspArgAspArgGlySerAsnAlaValIleHisTyrSerIleMetSer 440  
DB 220909 GTCAAGGCGCTGAGAGAGACAAAGGCGCAAGCGCTGAGTCAATATATGATCAATGAGC 220850  
QY 441 GlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
DB 220849 GGCACGCTCGGGGACATTTCTATCTGATCTCAGATCGAGACCTTGATGTGTGAGC 220790  
QY 461 ProLeuAspTyrGlnThrThrIleGlyGlnTyrThrLeuArgValArgAlaGlnAspGlyGly 480  
DB 220789 CCACCTCGACTATGAAACAACAAGAAATATACATCGCGAGATCCGGGGCAGAGATGTGGC 220730  
QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
DB 220729 CGGCTCCACTTTCCAAAGCTCTCGGCTTATGATCCGTCAGGTCTTGAATCAATGAC 220670  
QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSerValProLeu 520  
DB 220669 AACGCCCCCATCTTTGTAGAGACCCCTTTCAGGCGCACGGTCTTGAAGAGTACTTAA 220610  
QY 521 GlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540





QY	1102	-----	1102	QY	1102	-----	1102
Db	218389	TCACCTCATGGAACCCCTGTATCTCACACAGCCTTTGGCAGGTGAGAGGGCTACCAAG	218330	Db	217309	GCTTAGCCTTTGCTTCTCTACACAGGGGACACAGCCCAAGTCACAACTCTAGATAG	217250
QY	1102	-----	1102	QY	1102	-----	1102
Db	218329	ACACAGATTCCACCAAGGGGGTCAGAGACCTGCGCTCTCTCCCTCCCACTGTGTTC	218270	Db	217249	GGAAAGCCTGGCGCAGCTAGGGCCACTGTGGCTGCTTTCTGCCCCTGTGCCCTCC	217190
QY	1102	-----	1102	QY	1102	-----	1102
Db	218269	TCAGTAGCTGAGAACTGACTGTGACTTGTCTAGAGCACTCTTTCCCTCCAGGC	218210	Db	217189	CCATCAGCCTGACCTTAGATGATGACCAAGGCAACCCAAATTTCAGAGGCTGGG	217130
QY	1102	-----	1102	QY	1102	-----	1102
Db	218209	AGAAAGCCTGCTTGCTGCCCTTCAGAGCTCCAGGCAAGAAAGTACAGAGCTC	218150	Db	217129	TGAGGAGGACTGAGGCCCACTGCTTTGTCTTCTCTGTGGTTCTTTGTCTCTCT	217070
QY	1102	-----	1102	QY	1102	-----	1102
Db	218149	ATTGTCTGTCCAGACAGACCTCACTGCCCTGAGTTCTTGGTGGGGTCTGGCTTGGG	218090	Db	217069	TCCCTCCACCCCTGGCCCATGGGGACCTCTGCACTGCTGTGTCACTTCTCCAG	217010
QY	1102	-----	1102	QY	1102	-----	1102
Db	218089	GAAAGGCTGCACAAAGCCAGGGCAGATGACCTTTTGGGCTCCCCCAGCTGGCT	218030	Db	217009	CTTGTGTGCTTCTTCTTATAGCTTCTCTCACTGCTGGCTCTTGTCTTCCCTGTC	216950
QY	1102	-----	1102	QY	1102	-----	1102
Db	218029	TGTTCTTGGGGTGTAGTCTTCTCTGTGACCATGATGGCATGGCTTCAGGAAACAGA	217970	Db	216949	TCTATGTTTAGATCTGTCTTCTCTCTCACTGTGGCTGAGCTGCCCAAGCTGGG	216890
QY	1102	-----	1102	QY	1102	-----	1102
Db	217969	CACGGGGTTCAGAGTCACTAGGACTACTCTTTCAGGGGACAGACTCCAGGGCCTTG	217910	Db	216889	CTCATTGCTCTGACGCCCAATGTGCCGGCTGATCTCTTCAGACACACAGAGA	216830
QY	1102	-----	1102	QY	1102	-----	1102
Db	217909	ACTTGTGAGAGAGAACGATTTCTAGGGGGGTGGAGCTTGGCCAGAACCCAGCTTG	217850	Db	216829	CCTGCCAAGGCTGTCTCACTCTCCACAGACAGAGTCTTATGAAAGAACAAAGTTG	216770
QY	1102	-----	1102	QY	1102	-----	1102
Db	217849	GCATTCCTGACGTTGTCAAGGCTCAGCCCCCAGACAGAGAGATCGGGAGGGAAAG	217790	Db	216769	GGAGGCTGACTGTGGCTCTTCCCTGTGACATGGCTAGAGGCTTAGACCAAGTATAT	216710
QY	1102	-----	1102	QY	1102	-----	1102
Db	217789	CAGACTGGGTATATATCTATCTCTGTGACCTGTGAGATGAGAACAAAGGCGG	217730	Db	216709	TGCTGGGTCAAGCTTACCCAGCTCTCTTTTAAACAAATCAGGCCACTCTCTATTC	216650
QY	1102	-----	1102	QY	1102	-----	1102
Db	217729	GAGGGGGCAGACTCAGTAGCCAAAGACTTTCTGAGTCTGGAGTTGGGGCTGGAGAGG	217670	Db	216649	CCAACTGCTTTCGGAAGAAAGCAATGAGAGAGAGAGCAATGGTTGTTGCTG	216590
QY	1102	-----	1102	QY	1102	-----	1102
Db	217669	AACTGTATGTTTTTCAGGGGGTGAAGTATATGTGGCTTACTCCAGGCAAGTCTTACACATC	217610	Db	216589	GGATGGGACAGTAGACAGAGGGGACAGTGCACAGATGTGGCCAGATGGGGAGACAGC	216530
QY	1102	-----	1102	QY	1102	-----	1102
Db	217609	AGGAGGGGAGATGAGCTTCCGCTCTTCTGTCTATCTTCTTACGTGACAGACTT	217550	Db	216529	AAAGCATGCGGGCAGAGGGCAGATCAGGCTTGGCTGTGAACACATGCCGGAGTG	216470
QY	1102	-----	1102	QY	1102	-----	1102
Db	217549	CCGTATCCACCCCTCAGGTGGCTGCMAAGACCAAGGAGAAAGAGCTTCCAAAGGCTATG	217490	Db	216469	TGGCTTCCAGGACCAAGGATGGCAAGATTTCTTGTCTGTGGTTCAAGAGCACTGT	216410
QY	1102	-----	1102	QY	1102	-----	1102
Db	217489	GCTTGTCAAGCTGTAAACCACTGAAGCTCAGTTCCTCTCCAGAGAGAAAGTATAGA	217430	Db	216409	TTGTAGTCTCCCTGTGTCTTCTGTGACAGAGGAAATCCCATTCAGAAAACACTGGGC	216350
QY	1102	-----	1102	QY	1102	-----	1102
Db	217429	GCACTCTCACTGAGCTTGGCAGAGAGGCAAGAGCGGCTCTCCCTCCGCTTCCCA	217370	Db	216349	TGTAACTTCACTCTGTAGCCCAAGTCAAGGCTTCTGACAGAGGCAAGAGTGTGATC	216290
QY	1102	-----	1102	QY	1102	-----	1102
Db	217369	TCCACAGCTTCAGTGTGAATCCGGCTAGATTCTTAGGGCCCGGTGTGAGAGCAGCTCT	217310	Db	216289	ACCATGTGACCATGGCAGGCTGAGGGCTTAGAGGGCTGTGACTGTGTCTTGGTA	216230
QY	1102	-----	1102	QY	1102	-----	1102

D	b		216229	CCCTGGCAAGCAGGAGGACAGAGGTGACGACGTATGGCAGCGCTTGTGA	A	216170
O	y		1102	-----		1102
D	b		216169	CAGCAGATCACAAACAGAGAGTTGAAGGCAGGTAGCGCGCAAGCCAGGGCACA	T	216110
O	y		1102	-----		1102
D	b		216109	AGATGGGACTTGCTTCAGTATGCCAAGCCAATTGAAGTGTACATATACAA	T	216050
O	y		1102	-----		1102
D	b		216049	GTCCTATCGCTTTTCTACTTCTGCGTCAGCAGCATGCTTCAGTCCATCC	A	215990
O	y		1102	-----		1102
D	b		215989	GATGTATTTTGTCTTGATTGAGAGTTGCTCTCTCTCTCTCTCTCTCTCT	C	215930
O	y		1102	-----		1102
D	b		215929	TCT	C	215870
O	y		1102	-----		1102
D	b		215869	AGATGAATTTTTCTGTGTACTCTGTGTCTGTCTGTGAACCTGCTGTATG	T	215810
O	y		1102	-----		1102
D	b		215809	CCCTTGAATCAGAGATCTCTGCTCTGCTCTCTGACACTGTATTAAGTGA	C	215750
O	y		1102	-----		1102
D	b		215749	CACCAACACTAGTTCGCTGCCAGTTTAAAGTCTTTATGTAAAGGATTCAT	C	215690
O	y		1102	-----		1102
D	b		215689	AGTGTGAGTTTTTAATAAGGAATACCCCTATGACCCTTGAACCGTTCC	T	215630
O	y		1102	-----		1102
D	b		215629	TTGCTTCAGTTTTCATTTTAACTTAGTTCATTGCCGTGTGTATGAACAT	T	215570
O	y		1102	-----		1102
D	b		215569	GCATGTACTTCTGAGCCTAGGTGCGTGGGTGAAGTGGGCAACGCGATGT	A	215510
O	y		1102	-----		1102
D	b		215509	GTTTGTAGTGAAGCGGCTGCTAGGGGTATGCGAGCGACCTTTGGGTACA	C	215450
O	y		1102	-----		1102
D	b		215449	GAGCCAGTGAATGAATTCCTTAAGAGCCCTAGTGCAGAGAGAGACGCC	A	215390
O	y		1102	-----		1102
D	b		215389	GGGGCTCAGGGGTGAAGCTGCCCGTCTCTTCTATTCTAACAGATCCA	C	215330
O	y		1102	-----		1102
D	b		215329	CCCCACACCCTTTCTCAGACTTGGGGCACTGGGCGGTGTGACACTGG	C	215270
O	y		1102	-----		1102
D	b		215269	AGCTGATTTCTTGGGATGGGAGCCAGGCTGCCAGATTGATGAGAAGG	C	215210
O	y		1102	-----		1102
D	b		215209	TGTAAAGAGTGTGCACTGGGGCTAGCAGTTTGACCGAAGACAGTTATG	A	215150
O	y		1102	-----		1102

Dp	215149	CAAGGGAGAACAGAGTCAAGTGGAGAGACATGTTTCCAGAACGCAATTCGTGTCCACGCCAC	215090
Qy	1102	-----	1102
Dp	215089	CCGCCACAGCACGGTGAAGACGGGGTTATCTGGGGGTGAGGTGACGGGCGAGAGTC	215030
Qy	1102	-----	1102
Dp	215029	TCTCTGTGTTCTCTGGGCCCTCTAGAAAAGAAAGCAAAAGCCGTTTGGAGCAGG	214970
Qy	1102	-----	1102
Dp	214969	GAATTCAGAGCAGTCCGAGGGGACAGTGTATCTGAATAGGCTGAAAGGCAAGAACACT	214910
Qy	1102	-----	1102
Dp	214909	GCAGGCAAGTATGTTCAACGCTTATGCCAGAGAGTGGGGAAAAACAAAGAACGGTTT	214850
Qy	1102	-----	1102
Dp	214789	CAAGTGAACAGCTCAATGGCACTTTATGCATTAAGAAAAAGTTTCTGGGTACC	214730
Qy	1102	-----	1102
Dp	214729	TTGGCCGATATAGGGCTAGATTTTATAGCTGTGACCTTCTATCCGATGATCTCTTT	214670
Qy	1102	-----	1102
Dp	214669	ATGGTCTCCCTCTGGGTCTCTCCTCGAGCCCTCCCTCTTAGTACTCTTCTTC	214610
Qy	1102	-----	1102
Dp	214609	CTGGTCTCTCTCTCTCCGGGCCCTCCCTCTTGTACTCGGTTTCTAGTGTCTTACT	214550
Qy	1102	-----	1102
Dp	214549	CCTGATGCTCTCTCTCTGAGTCTCCCTTCTCAAGTACTTCTCCCTGGTGGTCC	214490
Qy	1102	-----	1102
Dp	214489	TCTGCAGACACTCCCTCTCTGTCTCGTTTTCTGTACTCTCTCTGACAAAGCTTAC	214430
Qy	1102	-----	1102
Dp	214429	TCCCTCTTCTCTTACGACAGGGTCTTGGCGATCTGCTTGCCTTACCCCAAGCTGA	214370
Qy	1103	-----SerApGjYALHsSerValThrAlaGlnCysAlaLe	1115
Dp	214369	TGTTGATTCGTGTGCTTCTCTTCAATAGGTGTCCACAGGTGACAGCCAGGCTCACT	214310
Qy	1115	UATGValThrLeIleThrAspGluMetLeuThrHisSerIleThrLeuAspGluAs	1135
Dp	214309	CCGTTGACCATATACACAGACGAGAGTCTCACACACAGATATACGCTGGCTTGAAGA	214250
Qy	1135	pMetSerProGluArgPheLeuSerProLeuGluGlyLeuPheIleGlnAlaValAlaI	1155
Dp	214249	CATGTCTCCAGAACGGTTTCTGTGCCACGTGCGAGGACTTTCATTCAGGCTGTGCAAG	214190
Qy	1155	ATHrLeuAlaThrProProAspHisValValValPheAsnValGlnArgAspThrAspAl	1175
Dp	214189	CACATTTGGCCACACCCCGAGATCACGTGGTGTCTTCAATGTGCAGAGGAGATATCTATGC	214130
Qy	1175	aProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGly	1195
Dp	214129	CCCAAGGGGCGCAATCTCTCAACTGAGCTGTGTCAAGTGGGCGAGCCCCCAAGACCCGGGG	214070
Qy	1195	YgIYProProPheLeuProSerGlnAspLeuGlnGlyLeuGlyYrLeuAsnArgSerLe	1215
Dp	214069	TGGGCGACCCCTTCTCTACGAGGATCTGCAGAGGCCCTGTATCTTCAACCGCAGCCT	214010

QY 1215 uleuthrala1ieseralagluarValLeupProphaeapapaam1leCysLeuarg1 1235  
DB 214009 GCTTACCCGCACTTCACTCAAGCGGTGCTCCCTTCGACGACACATTGGCTGGAGAA 213950  
QY 1235 uProCysgluAanTyMeArGcysValSerValLeuArGpheaSperSer1AproPh 1255  
DB 213949 ACCCTGCGAAGATTACATCGGTGTGTCCGTGCGCTTCGACTCCTCGCGCCCTT 213890  
QY 1255 el1e1aSerSerSerValLeuPheArGPro1eh1sP-rovalGlyglYLeuArGcysAr 1275  
DB 213889 TATCGCTTCCTTCGTGCTGTCTTCGCGCCATCCACCTGTCGGGGGTCTCGCTGTCCG 213830  
QY 1275 gCysPProGlyPheThrGlyAspTyrcysgluThruValAspLeuCysTySerAr 1295  
DB 213829 CTGCCCAACAGGCTTCAAGGCACTACTGCAAGTGAAGTGAAGTCTGTCTTACTCAAG 213770  
QY 1295 gProCysglYProh1sGlyArGcysArGserArGglYglYTyThrCysLeuCysAr 1315  
DB 213769 ACCTTGTCGACCCATGGGCACTGCCGACGTGAGAGGGGTGTATACCTGCTGTCCG 213710  
QY 1315 gAepGlyTyThr----- 1319  
DB 213709 CGATGCTACACGGGTAGTCTAGGCGAGGACAAATGGGGACCCCTGACAGGCTGACG 213650  
QY 1319 ----- 1319  
DB 213649 CTCTGGCCCAAGCACAAATCAGACTAATATAGGGGTGCTCATTTACTTCAAGGTATA 213590  
QY 1319 ----- 1319  
DB 213589 AACTAGATGCAACAGCGTTCTTAGATCTTACTGCTGTTTGAACAACCTCGCTTAGTA 213530  
QY 1319 ----- 1319  
DB 213529 AGTCAGAAATAGGAACCTTATGTTCACTGTGCAAGCCCAACCACTGCGACATGC 213470  
QY 1319 ----- 1319  
DB 213469 CGCGTCGCAACAGCCCAAGAAAGCTGTAAATTAGAGTCAGTCCGGGCACTGGGTTGT 213410  
QY 1319 ----- 1319  
DB 213409 GTTCAGCTTTTACCACTAATCTGTAAATAGAGGAACTTAGCTTCTTCCAC 213350  
QY 1319 ----- 1319  
DB 213349 AGTATTAAAGTAGATTAACCAATTTATGTAGTCAAAAAGTTGCTTGTAGAGACTCATG 213290  
QY 1319 ----- 1319  
DB 213289 TAAATAAATTTGATAGAGAGAAAGCTGGCTTATGTGTAGGTGTGGCTTTTGTAG 213230  
QY 1319 ----- 1319  
DB 213229 GAAGCTGAGATGAGGTGGCTTATCAATTAAAGGTGTGTGCTGAGCTTGAACAGAGC 213170  
QY 1319 ----- 1319  
DB 213169 CTTCAATATGCTGTGTGATCCCTTATTACTTGTAGCAAGCTTCCCAACCAAGAA 213110  
QY 1319 ----- 1319  
DB 213109 CTGAGAGAAAGCACTGAGAGCAGGGGTTTGAGTCAATGGAGTTCCAGAGACCTGAGCC 213050  
QY 1319 ----- 1319  
DB 213049 CGGTAGCCCTGCAAAATGGCTCATTTCTACCTCCAAATGTGTGAGGAGCTGAGGTCT 212990  
QY 1319 ----- 1319  
DB 212989 CGTTACACCTGCGCAACCTGTAGGCTGTGCTTCTCAAGCTCTTGTGTCTGTTTC 212930

QY 1319 ----- 1319  
DB 212929 TTCCAGACTCCAGGCTTAGGTCTTCAGCGCCAGCGCACTGCTCTGCAAGTGGC 212870  
QY 1319 ----- 1319  
DB 212869 TGGCAGACTCAGTGCCTCTGTGACACCTCTGTCTAGAACTTGCAGACTTACTCT 212810  
QY 1319 ----- 1319  
DB 212809 TCTGTGCTGCTGTTGTAGAGGTGGGGGTGAGGAGGAGGAGGAGTCTTTGTCC 212750  
QY 1319 ----- 1319  
DB 212749 AGAGAGATGGGAAAGGGGCTCCGAAAGTATGCGGAATTAACATTTCCATGGAACAAGA 212690  
QY 1319 ----- 1319  
DB 212689 GCCAGTCACTAGGCAAGGAACTAGAAATCTATCTAGAAATCAGAAATCGGAGGAGA 212630  
QY 1319 ----- 1319  
DB 212629 AGGAGTGGATTTTCCGTCTAATATCTGAATTTTCAAGCCCCGTGATTTCTGTCTCC 212570  
QY 1319 ----- 1319  
DB 212569 TCAGGCTTAGGCCCCCAGTCTTCAGGCTTGTTAGGCGGAGAGATCCTCAATTGAGA 212510  
QY 1319 ----- 1319  
DB 212509 CCCAATGTTCTGTCTTGGGAAAGCCCTGCTTACTTACACATGACACCACCAACTAT 212450  
QY 1319 ----- 1319  
DB 212449 AACTCAATATAGCTCGAAACCTTAGTCAATCTCTTAGGAAAGAAATCTGATCTTGACC 212390  
QY 1319 ----- 1319  
DB 212389 TGACAGAACAAAGCAGAGTGGGAAAGGGGAAAGAACGAGTGGGGCGGCGCTG 212330  
QY 1319 ----- 1319  
DB 212329 GAACCTGCAAGGCGCAGGGCTTAGGACAGGGTGACTGACAGGTGTGGGAAAGGCC 212270  
QY 1319 ----- 1319  
DB 212269 AAAGCTCACAGCAACATTTGTGAGCTGACAGAGAAATAGTGTGCAAGCCCAATAG 212210  
QY 1319 ----- 1319  
DB 212209 AGAGTGTCTGTCCATTAATTTCAAGATGTCAACAAAGAAAGTAGAAAGCATGGGA 212150  
QY 1319 ----- 1319  
DB 212149 GCCCGAGATGAGAGAGTCAACAGAGATTTTACAGGGAACAGTTAGACTTCTACT 212090  
QY 1319 ----- 1319  
DB 212089 ACTGATTAACCAAGAGAGACAGTTTAAACAAGAAAGTTTACTTGTCTCCGGTCTC 212030  
QY 1319 ----- 1319  
DB 212029 TGCAATTCAGTTCGTGTGCTGTCTTCGCTAATAGACAGACATCAGGGTGGCGAAG 211970  
QY 1319 ----- 1319  
DB 211969 CTGCTCACAGATGAGAAAGACCTGAGAACAAATGTAGCCCTGTATGGCGTACTCCA 211910  
QY 1319 ----- 1319  
DB 211909 GTGACTGCTTCTCAAGCTCTCTCTCCGATGCTTCAAGTGCCTCATTAACAGCCCC 211850  
QY 1319 ----- 1319

```

Db 211849 TCAGCAGACCAAGCCTTAGTATATAGCTGGGGGAGGGGGGCAATTATTAAC 211790
Qy 1319 ----- 1319
Db 211789 TGTAAACAGGGGACCTTCCAGCATCTCTGCAAGGCTCACCCTGCTTCCTCC 211730
Qy 1320 -GlyLysIleCysValIleSerAlaArgSerGlyValArgSerProGlyValCysLys 1339
Db 211729 AGGGAGACACTGTGAAGTGAAGTGGCCGCTCAGGCCGCTGTGATCCAGAGCTTCAGAA 211670
Qy 1339 nglyGlyThrCysValIleLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAs 1359
Db 211669 TGGGGTACCTGTGTCAACCTGTGTGGAGGCTTCAATGTGACTGCCATCCGGGGA 211610
Qy 1359 PheGlyLysProGlyCysGlnValThrThrArgSerPheProIleIleSerPheIleTh 1379
Db 211609 CTTTGAAGAGCCCTTCTGCAAGTGAACAAGAGCTTCCAGCGGCTCTTCATAC 211550
Qy 1379 rPheArgGlyLeuArgGlnArgPheIlePheThrLeuAlaLeuSer 1394
Db 211549 CTTCCGGGGGCTTGGCCAGGCTTCCACTTCACTTCCCTGCTGTGAGAGAGGCC 211490
Qy 1394 ----- 1394
Db 211489 CTGGGGTTCGAGGGCACTTGTAGAGCGTGGGGTACTCTGTTCAGTAGACGGGGTG 211430
Qy 1394 ----- 1394
Db 211429 CAGATCAGTAAAGAAAGTACCAAGAGAGTGGCTGACAGAGCACTTGATGCTTGCCA 211370
Qy 1394 ----- 1394
Db 211369 GCCGCGCGGCTCACTCAGAGATCTGGCTGCACTCTCCCATCTTGAGATGGGCG 211310
Qy 1394 ----- 1394
Db 211309 GCCAGAAATGCTGGGTTCTGCAATCTGACACAGCCTTAGCACAAGTGCAGTGTGC 211250
Qy 1394 ----- 1394
Db 211249 TGTGCTTAGATGCGCCAGAGTAGACACGCGGGTGGGTGCCCGTTCCTGGCCCGTT 211190
Qy 1395 ----- 1395
Db 211189 GACCATCCCCCTATACAGGTTTGTCTACCAAGAGAGGTGAAGGGCTACTGCTGAACGG 211130
Qy 1408 PheAlaThrLysGlyLysAspGlyLeuLeuLeuValLeuGln 1408
Db 1408 YArgPheAsnGlyLysIleAspPheValAlaLeuGlnValIleGlnValIleGlnLe 1428
Db 211129 GCGCTTCATAGAAAGCATGACTTGTAGCTCTGAGGTGATCCAGAGGAGTGCAGCT 211070
Qy 1428 uThrPheSerAla 1432
Db 211069 CACCTTCTCTGAGAGTAGAGCCAGCGCTCTGTCTTGGCCCACTCTTAAGATCTTAG 211010
Qy 1432 ----- 1432
Db 211009 TCTCCAACTCTGACTCAAGCATGCGCTCCCTCCAGTCAAGACTGAGCTGGGCT 210950
Qy 1432 ----- 1432
Db 210949 CCAATCTAGACATGCCCACTTCCCTGCACTCCACTCTTCTGTGACCCAGCTTAC 210890
Qy 1433 ----- 1433
Db 210889 CCGTCTTTTCTTAGGGGAATCCACACACAGGCTGTCTCACTTGTGCTTGAAGGGTCA 210830
Qy 1448 eArgPheGlyLysIleThrPheIleThrValGlnLeuLysIleValLeuLys 1462
Db 210829 GTGATGGCAAGTGCATACAGTACAGTGAAGTACTCAATTAAGTGGGTATGGGGGGA 210770
Qy 1462 ----- 1462

```

---

```

Db 210769 CAGAGAGCTAAGGGTCTGTCCGTGGCTGCTATTGTGAGTGTGGGAAGCTGTCA 210710
Qy 1462 ----- 1462
Db 210709 CAGCGCTGGGAAGAGAGAGAGTGTGCTTGCAGAGGCACTTCCATCCCTGTGTCT 210650
Qy 1462 ----- 1462
Db 210649 TGATGTGGGAATATGTGTAAAGCGAGTGTCTCCGTGGGTGTCTACTGCTTCTT 210590
Qy 1463 ----- 1463
Db 210589 CTCTGTCCAGACCGCTGTGTGGGTCAAGACAGACTTCCACAGGCCCATCTGAGCAGAAG 210530
Qy 1479 ValAlaValAlaThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerVal 1498
Db 210529 GTGCTGTGGTGTCCGTGAGATGGCTGTGACACAGGGTGTGCTGTGCTTCCGAGCTATG 210470
Qy 1499 LeuGlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyLysSerLys 1515
Db 210469 CTGGGCAACTACTCTGTGTCTGCTGCCAGGGCACCAGAGAGGCAAGAAATAGCAGAG 210410
Qy 1515 ----- 1515
Db 210409 ATGGGCGGGGAGAGGGGAGGTGACAGACTAAGGAGAGGTAGTGGGTCAACAG 210350
Qy 1515 ----- 1515
Db 210349 ATGGGGCTCCAGGTCAACAGAGATGGGGCTCCGAGATGAACTACATGCAATCCCTT 210290
Qy 1515 ----- 1515
Db 210289 CCGTCAATGTGGGAATTAGTCAAGTCTTGTGGGCTGTGAGGGGAAGGGAGCAGATC 210230
Qy 1515 ----- 1515
Db 210229 TGTCAATAGGGGTGCTGTCTCTCGCGCTTTTCTMAAAGGCTCATGAGCTCG 210170
Qy 1515 ----- 1515
Db 210169 GCTCAAGTCTCTCTTCTTGTGTCTAGTGGCAGGGTATGGGTAAGAAACCAAG 210110
Qy 1516 ----- 1516
Db 210109 ACTTAATCTGTGCTGTGCTTCTTCAATCTTGTGCTGCTCCAGGCTCTGAGCTGAC 210050
Qy 1520 rGlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMetArg 1540
Db 210049 AGGGCCCTGTGTCTGTGGTGGGTGCCAGATCTGCCGAGAGTTCCTGTCCGAATGCG 209990
Qy 1540 gGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAs 1560
Db 209989 GCACTTGTGGCTGTGATGAAGAACTCCAGGTGATAGCCGGCACTGCATGGCCGA 209930
Qy 1560 PheIleAlaAsnAsnGlyThrValProGly 1570
Db 209929 CTTCATGSCCAACAAATGGCACTGTGCTGG-TATGAGAACCTGGGGGTACGGCAAGGT 209871
Qy 1570 ----- 1570
Db 209870 GGGAGCCAAATGCCATGAGAGTGGTTGTGATGACACGGGAATGCCAGCTGCGGTT 209811
Qy 1570 ----- 1570
Db 209810 CTTCCGCTACCTGTGTGTTCCAGAGCGCTGGGGTCCCTGAGTCCCTCTCCGT 209751
Qy 1571 ----- 1571
Db 209750 TCTAAGCTGCCCAACAGAAAGAACTGTGTGACACCACTTGCATATATGTGTGAC 209691
Qy 1588 rCysValAsnGlnThrAspAlaPheSerCysGlyCysProLeuGlyPheGlyLysIle 1608
Db 209690 CTGTGTGAACAGTGGAGCATTCAGCTGCGAGTGTCCCTTAGGCTTCGGGGGCAAGAG 209631

```

```

QY      1608 rCyba1aGln----- 1611
Db      209630 CTGTGCCCCAGGGATGAGAGTGGGACAGCTGCCACAGATTGGGGCCCTGGGCTCTGTCAAGTTGTC 209571
QY      1611 ----- 1611
Db      209570 TGGGATTACGGGACAGGGCTTGGGGCAGGCTCTGGGAGGGCTCCCGGAGTCAAGGCTGTGG 209511
QY      1611 ----- 1611
Db      209510 GTGGAAAGACGCTGTCTGGGACAGTAAGAGAACTAGGACAGAGACAAAGACCTAGTTC 209451
QY      1611 ----- 1611
Db      209450 AAGAGCAGGCCCTGGGACAAATGTAGGGGGGCTGGAAACAAGTAAATGCTCCGGTGGGC 209391
QY      1611 ----- 1611
Db      209390 GGGGACATAGCCAGAACCCGGGTGTAGGGAGGCGAGGGTCCATCTCATGGAGCTGCCA 209331
QY      1612 ----- GluMetA 1614
Db      209330 CGCTTGTACTCCACTGCGCCCTGCTTAATCACTTTTCATCTGCTCCAGAAATGG 209271
QY      1614 1AAsnProGlnHisPheLeuGlySerSerLeuValAlaTPrHisGlyLeuSerLeuProI 1634
Db      209270 CCAATCCCAAGGGGTTTCTTGGGACAGGCTTGTGGCTGGGATGGCTCTCTCTGCCCCA 209211
QY      1634 1SerGlnProTPrpIyLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuL 1654
Db      209210 TCTCTCAAGCCCTGGACCTCAAGCTCATATGTCGACAGCCGAGGACAGATGGCGTCTTC 209151
QY      1654 euGlnAlaIleThrArgGlyAspSerThrIleThrLeu----- 1666
Db      209150 TCGAGGCCCTGACACAGGGGGCGAGACCAATCACCTTCAGAGGTGATGACAGGGCGCTGGG 209091
QY      1666 ----- 1666
Db      209090 AGGTGTGCAAGCCCGGCTTGGCTAACTTTGTAGAGACCTTCAGGGGGTTGGACCCAGGT 209031
QY      1666 ----- 1666
Db      209030 GGCTGGGACAGGGTTGGCAGAAAGCCAGACATGGCTGGGCCCATTCGGTGGGTCACTACT 208971
QY      1666 ----- 1666
Db      208970 GCCACCTATTGGGCTTAAAGAAATTAATCCCTCTGCTGCTCAGCGCGGACACTGG 208911
QY      1667 ----- GlnLeuArgGlnG 1671
Db      208910 AAGAGCCTACTGAGTGTGTCTACTAATTTTTTTCTTCCCTCTCGTACAGCTTCGAGGACG 208851
QY      1671 1HisIleMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnP 1691
Db      208850 GCCACCTAGTGTAAAGTGTGAGAGGACAGGGCTTCAGAGGCTCATCCCTGGCTGTGAGAC 208791
QY      1691 roGlyArgAlaAsnAspGlyAspTPrHisIleAlaGlnLeuAlaLeuGlyAlaSerGlyG 1711
Db      208790 CAGGCGGACGCAATGATGTGATCTGGCATCAAGCAGCTGTCACTGGAGACTAGGGGG 208731
QY      1711 1PProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuG 1731
Db      208730 GCCCTGGCCACGCAATCTGTCTCTTGTGACTATGGGACAAAGAAAGGAGAGGATTAATCTGG 208671
QY      1731 1PProArgLeuHisGlyLeuHisLeuSerPasnIleThrValGlyGlyIleProGlyProA 1751
Db      208670 GCCCTGGCTGACAGGGCTGCACTTGAAGCAATATGACAGTGGGGAGATTCCCGGGACAG 208611
QY      1751 1aGlyGlyValAlaAspGlyPheArgGlyCys----- 1761
Db      208610 CCAAGCAGTGTGGCCGTGTCTCCGGGGCTGTTTGACAGTGAATCTCTCCCGGACCTT 208551

```

---

```

QY      1762 ----- LeuG 1763
Db      208550 CCATCCACTTCTACTGAGACCCCTGTTCACATGAGGGCTAAATGCTGCTTATCCCTTC 208491
QY      1763 1nglyValArgValSerAspThrProGlnGlyValAlaAsnSerLeuAspProSerHisGlyG 1783
Db      208490 AGGGTGTAGGGGTAAAGCAGACACCCGAGGGGTGTAGACAGTGGATCCAGCCGCTGGGG 208431
QY      1783 1uSerIleAsnValGlnGlnGlySerLeuProAspProCysAspSerAsnProCysP 1803
Db      208430 AAGACATCAATGTGAGCCAGGCTGTAGCTGGCCAGATCCCTGTGACTCGAATCCATGTC 208371
QY      1803 roAlaAsnSerTyrCysSerAsnAspTPrAspSerTyrSerCysSerCysAsp----- 1820
Db      208370 CCACCACAGCTTACGACAGCAACGACTGGACAGTATTTCTTGTATGCTGTATCCAGGTA 208311
QY      1820 ----- 1820
Db      208310 AGCTGAGAGTCTGGGAAATGGGGCTGTGTGCCAGGGTCACTCCGTCTTAATATGTA 208251
QY      1820 ----- 1820
Db      208250 GCCAGTTACATCCAGGGTTATAGCAATCTGTGACAGGGTTCAAGCGGAAATTAAGGA 208191
QY      1820 ----- 1820
Db      208190 GTTGTAGATGAAAGAAATGCCAGTAAAGGGCGGGCTCGAACCTTTCCGATCCTTT 208131
QY      1821 --ProGlyTyrTyrGlyAspAsnGlyThrAsnValCysAspLeuAsnProCysGlnHisG 1840
Db      208130 GTCCAGGTACTTAATGTGTAACTGTATCAATATGTGTGTGACCTTAACCAATGACAGCACCC 208071
QY      1840 1uSerValCysThrArgGlyAspProSerAlaProHisGlyTyrThrCysGlnGlyCysProA 1860
Db      208070 AGCTGTGTGTACCCGAAACCCAGTGCACCCCAAGGCTTACATCTGCAAGTGTTCACCA 208011
QY      1860 snTyrLeuGlyProTyrCysGlnThrArgIle----- 1870
Db      208010 ATTACTTGGGCCCCATATGTGAGACAGGTA--GGTGGCGGGGCTTACTCTGTATGAC 207952
QY      1870 ----- 1870
Db      207951 TCAAGCCCGAAGTGCAGAGTGGGCTGAGCTGCAGCCCAATGAAGAATAATGTGGTT 207892
QY      1870 ----- 1870
Db      207891 GAACCAAGCGGAGAGATGCAAGTGTAGAGGGGCTCATAGCTGTTGGGTCCTTCT 207832
QY      1871 ----- AspGlnProCysAspProArgGlyTyrTPrGlyHisProThrCysGlyProCysA 1888
Db      207831 GCAGGATTTGACCAACCTTGGCCCCGCTGGCTGTGTGGGACACCCCAATGTGTCCATCA 207772
QY      1888 snCysAspValSerTyrGlyPheAspProAspCysAsnIleThrSerGlyGlnCysHisG 1908
Db      207771 ACTGCAATGTGACAAAGGCTTTTGAACCAAGCTGCACAAACAAACAGTGGCAGATGCCACT 207712
QY      1908 ySlyS----- 1909
Db      207711 GGAAGGTAGAGAACCCGAATGAGCCCGGCTGTGTACTCCTGTCTGTAGCTTGGGCC 207652
QY      1909 ----- 1909
Db      207651 TACAGGGAGACAGAGAGATGGGCTTCTTCAGAGAGCGAGACAGCATCTGTGCTT 207592
QY      1909 ----- 1909
Db      207591 TTGCTTATCAGGCAATGGAGAGTGTGTTGTTCCTTATATCCGCTGACAGGGAGATCCA 207532
QY      1909 ----- 1909
Db      207531 GTTCCAGTGTGAGTCAAGTGAAGCTGACATAATGTGTGGGCTGTTTGACAGCCATTCCTG 207472
QY      1909 ----- 1909

```

Db 207471 AGTATTGGCTTCTCACTGTAATGTCATTACTGCTGTTGAGAGTGTGAATCTCTCC 207412  
QY 1909 ----- 1909  
Db 207411 CACAAGACCCAGCATAGGGAGGCTCTCTCTGGGGAGGCCACAGCTTGAACCTTCGCT 207352  
QY 1910 ----- -GluAsnHisT 1913  
Db 207351 AGCATCCAGAAAGACATCTTGATCCAGTAGACCCCTTCTCCCTCCCTGAGAGAACTCACT 207292  
QY 1913 YTAAGProGlySerProThrCysLeuLeuCyAspCysIleProThrGlySerLeuS 1933  
Db 207391 ACCGACCCCGAGCGACCCCACTTGCCTCTGTGTGACTGTAACTCCCACTGGCTCTTGT 207232  
QY 1933 eATGValCyAspProGlyIleAspGlyIleCysProCysLeuProGlyValIleGlyAAG 1953  
Db 207331 CCGAGTCTGAGACCTTGAGAGAGCCAGTGCATGCAAGCTGAGAGTCATGGGCGTC 207172  
QY 1953 InCyAspArgCyAspAspProPheAlaGluValThrThrAsnGlyCyAsGlu----- 1970  
Db 207171 AGTGTGATCGCTGTGACAAACCTTTTGTGAGGTCAACCAACATGCTGTGAAGGTAGG 207112  
QY 1970 ----- 1970  
Db 207111 GGCTCTTAGCAGAGTAGGTTCCCTCCTCACAATATGCAAGAGCTTACACTCCAGAG 207052  
QY 1970 ----- 1970  
Db 207051 TTAGAAATGAGGCTGGCATTTGGGTGGGAGATGTGAGAGAAAGTTGGCGGCGATGATAG 206992  
QY 1971 ----- ValAsnTyrAspSerCyAsProArg 1979  
Db 206991 AACTGCTTACGATTAACCATGTTGTTCTTCCATATGAAATTATGACAGCTGCCACGG 206932  
QY 1979 laIleGluAlaGlyIleTTrProArgThrArgPheGlyLeuProAlaIleAlaProC 1999  
Db 206931 CCATAGAGGCTGGGATCTGTGTGGCGCCGACGCGTTTGGGCTACGTCCTCCCTC 206872  
QY 1999 YsProLysGlySerPhe----- 2004  
Db 206871 GCCCCAAGGGCTCTTGTGTAGTTTGAATTAGATTAATGATTAATGAACACTTGACTTT 206812  
QY 2004 ----- 2004  
Db 206811 ACTCTCAGAGGAATAAAGATAGCAGCTAGAGCCAGCTTGAAGCTCCCGGGGCCCACTT 206752  
QY 2004 ----- 2004  
Db 206751 TCTGGATTCTATCTCCAGAGACTAGTGTCTCACTTCACTGACATCAACCATCAA 206692  
QY 2004 ----- 2004  
Db 206691 CTTAGGCTGACAAACGTGCTTATATCCAGCCCAACAGAGCTGACGAAATCTCGGTG 206632  
QY 2004 ----- 2004  
Db 206631 TCCTGAACAGGGCTGGAGCTAGACTGTGTGATTGATGGGGTGCCTATCTTGAGCTAAC 206572  
QY 2004 ----- 2004  
Db 206571 TACTTAGTACTGAGTTAACCCACACTGAAGGAGGCGTGTGAGGATATGCGTGT 206512  
QY 2004 ----- 2004  
Db 206511 GGGTACTATCCCAAGAGATTCAAGATTAGTGTGACATGTTTCAATGGGAGGAGTAGCT 206452  
QY 2004 ----- 2004  
Db 206451 GAGAGTAAGCAGTAGCTGATGACAGGCGTAGTGAATATCTTGGGGCGGTGTGGAGA 206392  
QY 2004 ----- 2004

Db 206391 GGTACATAGGCTCCCTTTGGCCACTGATCATGATTTTGTGAGATGTTTGGCTCTGGAG 206332  
QY 2004 ----- 2004  
Db 206331 CTCTGTACACTGGGCTTCTCTGTCTCCCTAAGCCCGCTTCTTGCTTCTTCTCA 206272  
QY 2005 GlYThrAlaValArgHisCyAspGlyIleArgGlyTTrLeuProAsnLeuPheAsn 2024  
Db 206271 GGGACTGTGTGGCCCACTGTATGACACAGGGGCTGGCTCCGCAAACTTTCAC 206212  
QY 2025 CysThrSerIleThrPheSerGlyLeuLysGlyPheAla----- 2037  
Db 206211 TGCACTGAGTCACTTCTCAGAGCTTAAGGCTTTGT- AAGTTGACCCCTGTCTCTC 206153  
QY 2037 ----- 2037  
Db 206152 CTCTTCTCATACCTCTGTCTTTGAAATAAGAGCTCCCGGGTGGGATTAAGCTCAGT 206093  
QY 2037 ----- 2037  
Db 206092 GGTAGAGCGCTGTGCTAGAGAGCGCAAGGCCCTGGTTCGCTCCAGCTCCGGGAAAAA 206033  
QY 2037 ----- 2037  
Db 206032 AGAACCAAAAAAAAAAAAAAAAAAATAGAGCCGAATCTGTTCCGAGGCTCT 205973  
QY 2037 ----- 2037  
Db 205972 CTTACTGAGAGGCCACTTCTTCCCACTGATCTTGAAGCCCTTTTACTGTGAGCC 205913  
QY 2038 ----- GluArg 2039  
Db 205912 CAGTCCCAACCATCCGTGTGACCTCTGCACTTGTGCTCTCCATCCAGGCCAGCGG 205853  
QY 2040 LeuGlnArgAsnGlyIleuSerGlyLeuAspSerGlyArgSerGlnLeuAlaLeuLeu 2059  
Db 205852 CTACAGCGAAACCAATCAGGCGCTGACTCAGAGCGCTCCAGAGGCTGGCGCTGCTCTG 205793  
QY 2060 ArgAsnAlaThrGlnHisThrAlaGlyTTrPheGlySerAspValIleValAlaTyrGln 2079  
Db 205792 GGTATGCCACACAGCACCTCTGCTGCTTGGGAGTATGTCAAGGTGGCTTACAG 205733  
QY 2080 LeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThr 2099  
Db 205732 CTGGCCACAGACTGCTGGCTCATGAGAGTCCACAGGGGCTTTGGCTGTCCGCACA 205673  
QY 2100 GluAspValHisPheThrGlu----- 2106  
Db 205672 CAGGATGTGACTTCACTGAGTGGGCTTGAAGTAGCATGTGGGGCCAGTGAGAGA 205613  
QY 2106 ----- 2106  
Db 205612 GCTGTCTCTGAGATCTTGGGGAGATCTTGAAGGAGATCTTCTTACTGTAAAGAGC 205553  
QY 2106 ----- 2106  
Db 205552 TGCCCCACCCCTGTGAGAGGCTCCATGATAGTCTGTAAACAGAGCGCAACAGTTG 205493  
QY 2106 ----- 2106  
Db 205492 ATTAAGTATCACTCTCTGGGCTGGGATTTAGCTCAGTGTAGAGCGCTTAAGT 205433  
QY 2106 ----- 2106  
Db 205432 AAGCGAAGGCTGTGGTTCGTCCAGCTCCGAAAAAAGAAAAAAGAAAGTA 205373  
QY 2106 ----- 2106  
Db 205372 CTCTCTTACAGTGTCAAGAAATTAGTACTTGTGATTAAGGGCTGACCTGAGCA 205313  
QY 2106 ----- 2106  
Db 205312 GAGCCCACTTATTAACAAGAGCTAGTCTTGTGAGGAGCCAAAGAGGCAAAAACA 205253



QY 2106 ----- 2106  
 Db 205252 GAGCGACATGCAAGACACATGGGTGGCCCTGATTTAAAGTGAAGAGGTTTAA 205193  
 QY 2106 ----- 2106  
 Db 205192 GGCAGGATCCAGTCTGAGATGGAAGATATGACAGAGCTTCCAGGGGAGACTG 205133  
 QY 2106 ----- 2106  
 Db 205132 TGGGACTGCTAAGGGAGAGACCTAGTGGCGGGGTCTGAACCAAGGCAAGGCTGAACTT 205073  
 QY 2107 ----- AsnLeu 2109  
 Db 205072 GGTATGAGGTTGACCAAGCATATGCACTGTGACCATGCCATCTCCAGATCTGCTG 205013  
 QY 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisTyrGluLeuIleGln 2129  
 Db 205012 AGGGTGGGAGCGCCCTCTGATGCGGCATTAAGAGCACTGGGACCTGATCCAGCAG 204953  
 QY 2130 ThrGluGlyThrAlaTyrLeuLeuGlnHisTyrGluAlaTyrAlaSerAlaLeuAla 2149  
 Db 204952 ACAGAGGGTGGCACTGCTGGCTGCTCCAGCACTTAAGGCTTAATCCAGTGGCTGGCC 204893  
 QY 2150 GlnAsnMetArgHisThrTyrLeuSerProPheThrIleValThrProAsnIle 2167  
 Db 204892 CAGAAATATCGCGCATACTACTAGAGCCCTTCACTATCGTCAACCAACATTTGGTAG 204833  
 QY 2167 ----- 2167  
 Db 204832 GTGGTCTGGCGTGGGGCGGGGGTGTGAGAGTCTGCCCATTTCAAGGGCAGCCAG 204773  
 QY 2168 ----- ValIleSerValAlaArgLeu 2174  
 Db 204772 CGGTGCTGTGTAGTACTTGCACAAATATTTCTTCTAGTACTCTCTGTGGTGGCCCTG 204713  
 QY 2175 AspLysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgLysGln 2194  
 Db 204712 GATTAAGGGCAACTTCTCTGGACCAAGCTGCCCGTTATGAGGCGTTACGTGGGGAGCGC 204653  
 QY 2195 ProProAspLeuGluThrThrValIleLeuProGluSerValPheArg 2210  
 Db 204652 CTTCCAGATCTTGAGACCAAGCTATTTGGCCAGAGTCTGTCTTCAGAGGTCAAGCGGGA 204593  
 QY 2210 ----- 2210  
 Db 204592 AGTGAATGTGTAACGGCTGGGGGTGGGGCAACCATGCAAGTGGGAGGCGCACTGAGT 204533  
 QY 2211 ----- GluThrProProValValAlaArgProAlaGlyProGlyGluAla 2224  
 Db 204532 GTAGCAGCATTTACTTGCAGAGATGCTCCCATGTGTGAGATCTGAGGAGCACTGTGTAAGC 204473  
 QY 2224 AsnGluProGluGluLeuAlaArgArgGlnArgArgHisAspProLysLeuSerGlnIle 2244  
 Db 204472 CCAAGGACTGAGAGACTGGCCCGGGCGGCAAGGAGGCAACCAAGACTAAGTCAAGGGCCA 204413  
 QY 2244 ValAlaValAsnValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAs 2264  
 Db 204412 GGCAGTGGCCAGTGTATCATTTAACACACGCTGGCTGAGCTGTGCCCAACACTACGA 204353  
 QY 2264 ProAspLysArgSerLeu 2270  
 Db 204352 CCAAGATTAAGCCGTAGCTGAGTAAAGCGGCTACGACAGCTTAAGAGTGGGTTAAC 204293  
 QY 2270 ----- 2270  
 Db 204292 AAGGAAATGTGAAGATCTCCCAATGATTAAGGTTCTCTGCAAGCGGCTTTGGTACTGT 204233  
 QY 2271 ----- ArgValProLysArgProIleIleAsnThrProValValSerIle 2285  
 Db 204232 GTACCTTCTGTATCATAGAGTCCCAAGGCGCCAGTCAATCAACAACCTGTGTGAAGCAT 204173

QY 2285 eSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysProValIleThrValGlu 2305  
 Db 204112 CAGTGTCCAGATATATGAGAGCTCTCTCCACGGGCACTGGACCAAGCAGTCAAGTCCA 204113  
 QY 2305 nPheArgLeuLeuGluThrGluGluValGlyThrLysProIleCysValPheTyrAsnHisSe 2325  
 Db 204112 GTTCCGACTGTGAGAGCGAGGAGCGAACCAAGCCCATGTGTCTTCTGAAACATTTCC 204053  
 QY 2325 rIleLeu 2327  
 Db 204052 GATCTGTAGCCTGCTGCTTAACAACCTTAAGCCCATGACTTTGGACCAAGATCTCAGGC 203993  
 QY 2328 ----- ValSerGlyTyr 2331  
 Db 203992 TCCTGGCTGCTCACTCTTCTTCTCTCCGTAATCCGCTTCTCTCCGACAGGTCAGTGGCA 203933  
 QY 2331 hrGlyGlyTyrPheSerAlaArgLysGlyValValPheArgAsnGluSerHisValSerC 2351  
 Db 203932 CAGGTGGCTGGTCCGCGCAGAGGCTGCGAGGTGTCTTCCGTAAAGAGCCATGTCAAGCT 203873  
 QY 2351 YSGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArgArgLys 2369  
 Db 203872 GCGAGTGCACCAATATGACAAAGCTTTGCTGTACTATGAGATGTCTCCGACGAAGCTTG 203813  
 QY 2369 ----- 2369  
 Db 203812 GACTCCAGGGGTAGCCCAAGAGCAGCGGGGTGGGACCAATGTGAGAGGCTCAGGCGCTG 203753  
 QY 2370 ----- AsnGlyGluIleLeuProLeuLysThrL 2379  
 Db 203752 CTCTTTTGTGACCCCAACACTATCCCTCAGAAATGTGAGATTTTGCACCTGAAGACCC 203693  
 QY 2379 eutThrTyrValAlaLeuGlyValAlaThrLeuAlaAlaLeuLeuThrPhePhePheLeuT 2399  
 Db 203692 TGAGGTATGTGGCCCTTGGATCACTTAAGTGTGCTGATATATCACTTCTCTTCCCA 203633  
 QY 2399 hrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaL 2419  
 Db 203632 CCCTCCTTTCGAGCCCTTCTGCTCCCAACGAGTGCATTCGACGCAACCTCAAGCTGCC 203573  
 QY 2419 eugLysLeuAlaGluLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuPro 2436  
 Db 203572 TGGGCTGGCCAGCTGTCTTCTCTGGGAAATCAACAGGCTGAGCTT - CCTGTAAAGA 203514  
 QY 2436 ----- 2436  
 Db 203513 TACTTCTCTCCAGAGACTTCCCAACTTCAAGCCCAAGGCTCATGTGTCCCT 203454  
 QY 2436 ----- 2436  
 Db 203453 CTATGACCTTGTCCATCCCTGCTCCCAAGCAATCAATAGTAAGGAGCCCATTCAGT 203394  
 QY 2436 ----- 2436  
 Db 203393 ACTGCGCAGAGTTGGCCCTTCTCTACACACACAGGCTTCCCACTCCCTGGGGGAGC 203334  
 QY 2437 ----- PheAlaCysThrVal 2441  
 Db 203333 TCCTTGTATGACTGTGTGGCTCTGTATCATATGTCCTGGCCCTCAGTTTCTTGACAG 203274  
 QY 2441 alIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeuGluVal 2461  
 Db 203273 TCATTTGCATCTGTGACTGTCTGTACCTTTCACCTTTCCTGTGGCTCTGTGTGAGG 203214  
 QY 2461 lAlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgP 2481  
 Db 203213 CCTTAACCTGTACCGGGCGGTCAAGAGTGCAGAGCTCAAGGCCCACTGCGTT 203154  
 QY 2481 heTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr 2494  
 Db 203153 TCTACTACATCTGGGCTGGGGCGTTCCTGCTTCAATCAACAGTATCTCGCTCTTTC 203094  
 QY 2494 ----- 2494

```

Db 203093 TGAACCTAGAGCTCCCTGTTGCTAGAGCTGAGGCTACCTTCATGATGATACCCCAAC 203034
Qy 2494 ----- 2494
Db 203033 CATAGAAATTCGAGAAATCCCTTCCCTTACTTGGACAGCCTGGATTAGGTGTGAGG 202974
Qy 2494 ----- 2494
Db 202973 GCTAGAGCTTCTCCTTGGAGAGGAGTGTGACAGACGCTCTCTGTATGTCTCCCACTG 202914
Qy 2494 ----- 2494
Db 202913 GCTCTGTGAACCTTGTATGATCATATGCTCCACCTGAGCACTGTCTGTGCT 202854
Qy 2495 -----GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAsp 2510
Db 202853 GTCTCTGTGCTCCAGGCTGTGGCTGTGGGCTGTGATCCCAAGGCTATGAAACCTGACT 202794
Qy 2510 heCyATrPLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPhe 2530
Db 202793 TCTGCTGGCTCTTCATCTACGATACGATACGCTCATCTGAGTTTGTGTGACAGTGGCTTTG 202734
Qy 2530 laValSer----- 2532
Db 202733 CTGTTTCGGTAAAGTCTTGAAGCAATTAGGATGACAGATGCGAGGCTCTACAGAT 202674
Qy 2532 ----- 2532
Db 202673 TCCTCTTTAGGAAGCTAGAGCTCTGAGGCTCCCTGGCAAGAAAGTCTCTGCTCC 202614
Qy 2533 -----MetSerValPheLeuTyrIleLeuAlaAla 2543
Db 202613 TCTCTGCTCTCATATGCTCCCACTGACATGATGATGCTCTTCCTGATATCTCTGTGGGCC 202554
Qy 2543 hGAlaSerCyAlaIleAlaGlnArgGlnGlyPheGlnuLyArgGlyProVal----- 2559
Db 202553 GAGCTCTCGTGTGCTGCCAAGCGTCAAGGGCTTGCAGAAAGGCTCTGTGATGATCTGG 202494
Qy 2559 ----- 2559
Db 202493 CAGTGGCTGGGACAGGGCGGGTGTGTATAGAGATCATGGCGGCTGTGTATTA 202434
Qy 2560 -----SerGlyLeuGlnProSer 2565
Db 202433 GAGAGCATCATGCGGAACTCAATATGCTGCTCCCTCCAGCTCAAGGCTGTGCTCTCC 202374
Qy 2566 PheAlaValLeuLeuLeuSerIleThrTrpLeuLeuAlaLeuLeuSerValIleAsnSer 2585
Db 202373 TTCACTGTCTCTGCTGTGATGATGACATGCTGTGGCTGTGCTGTCTGTCAACAGC 202314
Qy 2586 AspThrLeuLeuPheIleTyrIlePheAlaThrCyAsnCyAla----- 2600
Db 202313 GACACTCTCTCTTCACTACCTCTTCCGGGCTGTGAATTGTGT-CCAGGTACTTAGCCC 202255
Qy 2600 ----- 2600
Db 202254 CGCCCTTACGATGAGAGAGATGCTGCGGGTGGGATGTGTGCTCAACCATGCTGTCT 202195
Qy 2601 -----GlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGluValArgLys 2618
Db 202194 GCCTGCGAAGGCGCCCTCAATCTTCTCTCACTGATGTGTCTCAAGAGAGGATCGGAAA 202135
Qy 2619 AlaLeuLysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSer 2638
Db 202134 GCACTCAAGTTGCGCTGACAGCGGAAGCCAGCGCTGACCTGTGCTTGAACCTAAGTGT 202075
Qy 2639 ThrLeuThrSer----- 2642
Db 202074 ACCCTGAACCTC-GGTGAAGGAGGCGGGGTCTGAGCTCAAGAGCGCTTGTGTAGAAGGG 202016
Qy 2642 ----- 2642

```

---

```

Db 202015 AAGCAGGCCCCAGAAACCTGTTGTCTTGTACTTTTAACTGCGCATCCTTTTCC 201956
Qy 2642 ----- 2642
Db 201955 CATCTCTCATCAGAGAGAGGGGTGAAGTGAAGGGGTGTCTCCTTGGAGAAATGCTA 201896
Qy 2642 ----- 2642
Db 201895 TCACGTGGTAAAGCTTTTGACTTGGGGGAGATGCAAGCAAGATTTGGTTGTCAAGTGG 201836
Qy 2642 ----- 2642
Db 201835 ACACATGTGTAGAAATTCCTCATGAAATTTGATATTAAGTTTCTGTCTCTAGAAATAT 201776
Qy 2642 ----- 2642
Db 201775 CTCAGAGACAAATTCAAAGTCCATTAATAAAAAAAAAAAAAAAAAACAGCAAGCCTTATA 201716
Qy 2642 ----- 2642
Db 201715 CAAGTCCGGGGGTTTGTCTGTATTAGTCCACAGCTAGCGTATAGCTGTGCCACTG 201656
Qy 2642 ----- 2642
Db 201655 CTGTGCCCTGTGTGAAGCAACAGAAAGGCTTCTCGTTAGCCCTGTGGCATGTGGG 201596
Qy 2642 ----- 2642
Db 201595 AGTTAGGGCTTGGGTAAAGTTCTGTGGGCAACCCCTCCACTCACTCACTCATCTGGCAT 201536
Qy 2643 -----SerTyrAsnCyAspProSerProTyrAlaAspGlyArgLeuTyr 2656
Db 201535 GTCTGTCTCTTGCCTTAAGTCTTATTAAGTCCCAAGCCCTTATGAGATGGAGGCTGTAC 201476
Qy 2657 GlnProTyrGlyAspSerIleGlySerLeuIleSerThrSerArgSerGlyLysSerGln 2676
Db 201475 CAGCTTACGAGAGATTCGGGCTCTCTTGCACAGTGCACGCCCTCAGGCAAGTGTAG 201416
Qy 2677 ProSerTyrIleProPheLeu----- 2684
Db 201415 CCAAGCTACATCCCTCTTCTGTGATGATGACCCCGGAAATGGAGAGACTGGGAAAGAG 201356
Qy 2684 ----- 2684
Db 201355 GAAAGACTGATGCTAGAACCAAGGCGCGCAGCTGTAGAGCGAGTGTCTACACTG 201296
Qy 2684 ----- 2684
Db 201295 CGGCCAACAGATGCACAGTTTAGAAGCATAGAGGACAAAGGTAGGGCTGGGAATCT 201236
Qy 2685 -----ArgGluIleSerAl 2689
Db 201235 CACTAGAGAGGCGACATGACACCTTACAGCATAGTCTTGTCTTTTATGAGGAGGTCCAC 201176
Qy 2689 AlaLeuAspProGlyGlnGlyProProGlyLysLeuGlyAspProGlySerLeuPheLeuGlu 2709
Db 201175 GCTGAACCTGTGGCAGAGTTCCCTCTGTGGGAGACCAAGTGGCTATTCATGAAAGG 201116
Qy 2709 YGluAspGlnGlnHis----- 2714
Db 201115 CCAAGCCAGCAACAGGTGAGAGACAGAGGCTTAGATGCTTAGCAGGGCAGGGGCTGTG 201056
Qy 2715 -----A 2715
Db 201055 CTTTTCGTGAAGGGGGGTGGGGCAGGTGTGTATTCCTCCTCCTGACCTCCAG 200996
Qy 2715 AspProAspThrAspSerAspSerAspLeuSerIleGluAspAspGlnSerGlySerTyr 2735
Db 200995 ATCCCGACACCGACCTGACAGTGAACCTGTCTCCGTGAAGATGACCAAGTGTCTTACG 200936
Qy 2735 laSerThrIleSerSerAspSerGluGluGluGluGluGluGluGluGluGluGluGluGlu 2755
Db 200935 CTCTTACCACCTCATTCAGACAGC-----GAGGAAGAGAGAGAGGCGCGCT 200888

```

```

QY 2755 heptgylgylnglYTPASPserleuenglYProglYalaglYargleuProleuH 2775
Db 200887 TCCGCGCAGCAGGAGGCTGGGAGACGCTGCTGGGCTCTGGAGGCTGAGACCTGCCCTTC 200828
QY 2775 isgertHrProlys----- 2779
Db 200827 ACAGTACCCCA- GGGTGGGCCAGCAGCAGGCTGAGCCATTGGGGGGCAACGAGAG 200769
QY 2779 ----- 2779
Db 200768 CTTAAGAGCTTGGGCTTTGGCTTACCTGCTCTGAGACTATAGGCTTCC 200709
QY 2780 -----AspYgylYProglYProglYalProglYaspPheglY 2796
Db 200708 TTTCTTTCAGACAGATGGGGCTCCAGGCTCTGGGAAAGTCCCTTGGCCAGAGACTTTGGGA 200649
QY 2796 htrhralYargleuSerSerYAsnglYalProglYgluYgluYargleuYAsnglY 2816
Db 200648 CCAACAAAGAGAAATGATGTAGTGGGCCCTTGAAGGCGGCCACGAGAGAAATGGAG 200589
QY 2816 spAlaIeuserArGlYglYSerleuGlYProleuProglYserSerAlaGlYProHle 2836
Db 200588 ATGCTCTAAGCTCGGAGAGGCTCTGGGACCCCTTCCGGCCCTTCTACCACTCA 200529
QY 2836 YsglYlYle----- 2838
Db 200528 AAGGTTA-GTGAAGAGTCTCTCCATCTGCTTCACTGCGATCCTTGGTGGAAATTTT 200470
QY 2839 -----LeuYglYglYglY 2844
Db 200469 AGAATGCTCATCTCATGTGTAACCAACCTTTACCAAGCATCTCCAAAGAGTGTG 200410
QY 2844 euProThrIleSerGlYlYserSerleuYargleuProleuGlYnglYserThglY 2864
Db 200409 TGGCCACATCAGCGAAAGATGAGTCTCTAAGTTGGCCCTGGAGCAGGAGCAGGGT 200350
QY 2864 exSerArGlYserSerAlaSerGlYglYserArGlYglYProProProArGlYPro 2884
Db 200349 CTTCCGGGGGCTCCACCGCAGTGAAGGAGGAGCAGAGAAAGGACCTCTCCACGCCAC 200230
QY 2884 roArGlYserleuGlYnglYgluYgluYglYalMetProHlealMetSerleuY 2904
Db 200289 CGGCGCAGTCTCTCCAGAAACAGCTGATGAGGCTCATGCCATGAGCATCAAG 200230
QY 2904 laGlYThrValaSPglYaspSerSerGlYserGlY 2915
Db 200229 CAGGCACTGTGATGAGACTTTTGGCTCCAG 200195

RESULT 13
AC095463/c 235327 bp DNA linear HTG 09-MAY-2003
LOCUS AC095463
DEFINITION Rattus norvegicus clone CH230-7H8, WORKING DRAFT SEQUENCE.
ACCESSION AC095463
VERSION AC095463.6 GI:30467711
KEYWORDS HTG; HTS_PHASE2; HTS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 235327)
REFERENCE
  Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
  Allen, C., Allen, H., Albrooke, S., Amin, A., Anguiano, D.,
  Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
  Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F.,
  Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
  Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
  Chacko, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
  Chao, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
  Cleveland, C., Cockrell, R., Cox, C., Coyte, M., Cree, A., D'Souza, L.,
  Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
  Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

```

```

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Esposito, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, R., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hune, J., Idlibird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,
Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahdardine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundada, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackeme, O., Okwom, G., Olampunaggon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Prims, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartbeyn, A., Slason, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, G., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 235327)
REFERENCE
  Worley, K.C.
  Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 235327)
  Rat Genome Sequencing Consortium.
  Direct Submission
  Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On May 9, 2003 this sequence version replaced gi:24941094.
  The sequence in this assembly is a combination of BAC based reads
  and whole genome shotgun sequencing reads assembled using Atlas
  (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
  in the feature table below represents a scaffold in the Atlas
  assembly (a 'contig-scaffold'). Within each contig-scaffold,
  individual sequence contigs are ordered and oriented, and separated
  by sized gaps filled with Ns to the estimated size. The sequence
  may extend beyond the ends of the clone and there may be sequence
  contigs within a contig-scaffold that consist entirely of whole
  genome shotgun sequence reads. Both end sequences and whole genome
  shotgun sequence only contigs will be indicated in the feature
  table.
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: GCJO
  Center clone name: CH230-7H8

```

## ----- Summary Statistics

Assembly Program: Atlas;  
 Consensus quality: 224990 bases at least Q40  
 Consensus quality: 226607 bases at least Q30  
 Consensus quality: 227843 bases at least Q20  
 Estimated insert size: 236896; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.bjesc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.bjesc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 235327: contig of 235327 bp in length.  
 \* Location/Qualifiers  
 1..235327

FEATURES  
 source  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-7H8"

BASE COUNT 58928 a 53675 c 55037 g 61147 t 6540 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 235327  
 Score: 12001.00 Matches: 2789  
 Percent Similarity: 38.64% Conservative: 47  
 Best Local Similarity: 38.00% Mismatches: 75  
 Query Match: 77.20% Indels: 4436  
 DB: 2 Gaps: 33

US-09-916-849a-3 (1-2923) x AC095463 (1-235327)

QY 1 MetatgserProAlaThrGlyValProleuProThrProProProleuLeu 20  
 DB 224814 ATGGGAGACCGGGCTGCGACGCCGCCCTCCCAAGCCACCTGCTGCTGCTG 224755  
 QY 21 LeuleuLeuLeuProProProleuLeuGlyAaPglValGlyProCyAaArgSerleu 40  
 DB 224754 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 224695  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAaCyAaPglValGlyProCyAaArgSerleu 60  
 DB 224694 GGGTCGGGGGACCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 224635  
 QY 61 SerAlaSerAsnLeuThrLeuThrLeuThrSerArgCyAaArgAaPglValGlyThrGlyLeuThr 80  
 DB 224634 TCTGCTTGGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 224575  
 QY 81 GlyHisLeuValProHisHisAaPglValAaArgValAaProCyAaArgSerleuAaHis 100  
 DB 224574 GGGCCACCTGGTGGCCCATCAAGATGGCTGAGGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 224515  
 QY 101 IleProleuProProAlaProGluGlyCyAaProTTPSerCyAaArgLeuLeuGlyValLeuGly 120  
 DB 224514 ATCCCTCTCCGCAATCTCTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 224455  
 QY 121 GlyHisLeuSerProGluGlyValLeuThrLeuProGluGluHisAaProCyAaLeuValAa 140  
 DB 224454 GGACACCTTTTCCCAAGAGGACGCTGACCTGCTCAAGAAACACCTTGTAAAGGCC 224395  
 QY 141 ProAaGleuAaArgCyAaGlnSerCyAaValLeuAaGlnAaProGlyValAaArgValAaGlyVal 160  
 DB 224394 CCACGCTCAGATGTCAGCTTGGCAAGGCTGCAAGGCGCCCAAGGCTCAAGGCTGAGGGA 224335  
 QY 161 ArgSerProGluGluSerLeuGlyValAaArgValAaArgAaValAaThrAlaProGln 180  
 DB 161 ArgSerProGluGluSerLeuGlyValAaArgValAaArgAaValAaThrAlaProGln 180

DB 224334 GGATCAGACAGAGATCTATGGTGGCGGACGAAAGAAAGTGTATATACCTCCAGC 224275  
 QY 161 PheGlnProProSerGlyGlnAlaThrValProGluAaGlnProAlaGlyThrProVal 200  
 DB 224274 TTCCAGCTCCCAAGCTTACAGACCAAGTGGCCGAGAAACAGGCTGCTGTAACCTCTT 224215  
 QY 201 AlaSerLeuAaGlnAaLeuAaPglValGlyValAaGlyAaArgSerleuAaThrMet 220  
 DB 224214 GCTCTTAAAGGACATTTGATCCCGATGAGGGTGAAGGGTCCAGCTTGAATACCAATG 224155  
 QY 221 AsnAlaLeuPheAaSerAaSerAaGlnPhePheSerLeuAaProValThrGlyValAa 240  
 DB 224154 GATGGCTCTTGTGATAGCGCTGCAATCAATTCCTCTTGGAGCCCAATCACTGGTCT 224095  
 QY 241 ValThrThrAlaGluGluLeuAaPglValThrValAaArgValAaPheAaValThr 260  
 DB 224094 GTACACAGCTTGAAGACCTGATGCGGAGACCAAGACCAACCAATGTTCAAGGCTCAC 224035  
 QY 261 AlaGlnAaPheAaGlyMetProAaArgSerAaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 224034 GCACAGATCATGGATAGCCCGACGAGAGCTTGGCCACATTAATCTTATGTAAT 223975  
 QY 281 AsnThrAaAaPheAaAaPglValPheGluGlnGlnGlyValGlySerProSerGluValAa 300  
 DB 223974 GACACCAATGATCAAGACCCCTGTTTGAAGCAGCAAGATCAAGAGAGCCTCAGGGAG 223915  
 QY 301 AsnLeuGluValGlyValGluValLeuThrValAaArgAlaThrAaPglValAaPheAa 320  
 DB 223914 AACCTGAGGTTGGCTATGAGGCTTCAAGTCAAGGACCAAGGACCAAGGAGGCTCTCT 223855  
 QY 321 AsnAlaAaAaLeuValAaArgLeuLeuGluGlySerGlySerProSerGluValAaPhe 340  
 DB 223854 AATGCCAATATTTGTAACCGCTGCTGAGAGGGCTGAGAGGACCCCTCAAGAGTCTTT 223795  
 QY 341 GluIleAaPheAaAaSerGlyValAaIleArgThrArgGlyProValAaPheGluGluVal 360  
 DB 223794 GAGATGATCTCTGCTGCTGAGGATATCCGAACCGCGGCCCCCTGATGATGGGAAGATG 223735  
 QY 361 GluSerGlyGlnLeuThrValGluAaSerAaPglValAaArgAaPglValProAaSer 380  
 DB 223734 GAATCTTCAAAATTTGATAGTGAAGGCAAGTGAACAGGCTCGGACCCCAAGGCCAGAGC 223675  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAaPheAaAaAaAaAaAaAaAaAaAaAaAa 400  
 DB 223674 TCCACAGCTGCTGCTCTCTGCTGCTGAGATGATATGACCAAGCCCAAGTTTAC 223615  
 QY 401 GluValAaArgGlyValAaGlnValAaArgValAaPheValThrProGlyValAaProValLeuAa 420  
 DB 223614 GAGAAAGGCTAAGTGTGAGGAGGAGGAGATGTGACCCCAAGAGGCCCGCTTCTCGA 223555  
 QY 421 ValThrAlaSerAaPheAaAaPglValGlySerAaAlaValAaHisGlySerIleMetSer 440  
 DB 223554 GTCAAGGCTCTGACAGACAGACAGAGGACCAAGCCCTGCTGCTATATGATATAGC 223495  
 QY 441 GlyAaAlaArgGlyGlnPheValLeuAaPheAaGlnThrGlyValAaLeuAaPheValAaSer 460  
 DB 223494 GGCAACCTCGGGGACAGTCTATCTGATGCTCAAGCTGAGGCCCTGATGTGTGAGC 223435  
 QY 461 ProleuAaPglValGluThrValAaGlyValAaArgValAaGlnAaPheGlyValAa 480  
 DB 223434 CCACTGCACTATGAGAACCAAGAAATATACCTGAGATCGGGGCCAGAGATGTGGC 223375  
 QY 481 ArgProProleuSerAaValAaSerGlyValAaValAaThrValAaGlnValAaAaPheAaAaPhe 500  
 DB 223374 CGGCTCACTTTCCCAAGCTCTCGGCTGATGATACCTGAGGCTTCAAGCAATCAATGAC 223315  
 QY 501 AsnAlaProIlePheValAaSerThrProPheGlnAlaThrValLeuGluSerValProleu 520  
 DB 223314 AACGCCCCCATCTTTTGCAGACACCCCTTTCCAGGCCACGCTCTCAAGAGATGTATCACTTTA 223255  
 QY 521 GlyThrLeuValLeuHisValGlnAlaAaLeuAaAaPheAaGlyAaPheAaAaAaAaAaAaAa 540  
 DB 223254 GGCTACTGCTGCTGATGCTCAAGGATGATGCTGATGCTCGAGAGAACGCGGCTCT 223195

Qy	541	GIuTyArzArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly	560	Qy	901	AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProVal	920
Db	223194	GAATACAGCTGGCTGGGGTGGGCAAGACTTCCCTTCAACATTAAACAAGGCAAGGC	223135	Db	222114	GCCGCAACACCAATGAAGTACAGTTACTGTATCGATGTGAATGACAATCCCGCTC	222055
Qy	561	TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal	580	Qy	921	PheGluGlnAspGluPheAspValPheValGluGluIleAsnSerProIleGlyLeuAlaVal	940
Db	223134	TGGATTCGGTGGCCCGGAGTGGACCGGGAAGAGTTGATTCTACAGCTTCGAGTA	223075	Db	222054	TTTGAAACGAGTGAAGTTGATGTTATTTGTGGAAGAACACCCCATGTGACTGGCTGTG	221995
Qy	581	AlaAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThrVal	600	Qy	941	AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle	960
Db	223074	GAAGCAAGGACCAAGGACCCCAAGACTACCGCTCAGCCAGTTCAGTGAACCAATC	223015	Db	221994	GCCCGGCTCAACGCCACCAAGCCCAAGAGGACCAATGACAGATATGACCAATTA	221935
Qy	601	LeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620	Qy	961	ValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThrAla	980
Db	223014	CTGATGTCAATGACAAACAACCTTTACGACCGGAGTGAACGGTGGCGCTCAAC	222955	Db	221934	GTTGAGGGCAATATCTCGAGGTCTTTCAAGCTGACATCTTCGGCGGAGCTAACCGCC	221875
Qy	621	GluAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis	640	Qy	981	LeuValAspLeuAspTyrGlyLeuAspArgProGluTyrValIleLeuValIleGlnAlaThrSer	1000
Db	222954	GAGGACCGCGCTGTGGCACCACTGTGTGACCGTGTCAAGCCCTGATGAGATGCTCAC	222895	Db	221874	CTGTGAATTTGGACTTACGAGAACCGACCTGAATATATCTGTGTATCAGGCTACGCT	221815
Qy	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660	Qy	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro	1020
Db	222894	AGGTGATTAACCTACAGATACCAAGGCAACCCGCAACCGCTTCTCTATCACACG	222835	Db	221814	GCTCCCTGTGGAGGAGGCTACTGTCTCATGTCCGCTCTCGACCGCAATGATACCCG	221755
Qy	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleLeuGluArgGln	680	Qy	1021	ProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040
Db	222834	CAAAAGGGGGTGGCTGGTCTCCCTGGCTTACCCCTGAGCTACAAACTCGAGCGGCAAG	222775	Db	221754	CCAAGTCTGGGCAACTTTGAGATCTTTTCAACAACATATGTCACCAACCGCTCGAGACG	221695
Qy	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700	Qy	1041	PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu	1060
Db	222774	TATGTCTGGCCCTGACCTGCTGATGTGGCAAAAGGACAGACACACTCAAGTCCGTGGTG	222715	Db	221694	TTCCCTGGGGGTGCTATAGGCGCGGTGCTGCCATGACCCCGAATCTCAGACGCTGTG	221635
Qy	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720	Qy	1061	ThrTyrSerPheGluArgGlyAsnGlyLeuLeuSerLeuValIleLeuAsnAlaSerThrGly	1080
Db	222714	AATGTCACTGATGCCAACCAACCTTCCTCCCTTCCAGAGCTCCCACTATACGAGTAAT	222655	Db	221634	ACGTACACTTTGACCGGAGGAATGACTCAAGCCCTGCTCTACTCAATGCTCCACCGGT	221575
Qy	721	ValAsnGluAspArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAsp	740	Qy	1081	GluLeuGlyLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal	1100
Db	222654	GTTAATGAAGACCGGCAAGGACCAACAGTGTCTGATGCTGCTACAGTCAAGCAGAGAC	222595	Db	221574	GAGCTGAAGCTGAGCCGGGCACTGGACAACAACCGGCCCTGGAAGCATCATGTAGTGTG	221515
Qy	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760	Qy	1101	LeuVal	1102
Db	222594	ACAGGGAGAGATGCCGAATCACTTATATGAGAGACAGCATCCCTCAAGTCCGCAATC	222535	Db	221514	CTGGTGTGAGTGAAGAAAGATGCCAGTGAATGTGGGTGGGAATATGCTTTAGGAGG	221455
Qy	761	AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGlyLeuAspGlnValSer	780	Qy	1102	TCCTTGAAGCACTGGAGGTGTGGCTTCCAAAGAGGGGCTGTGACCAAGCTCTCCAC	221395
Db	222534	GATGCAAGACCTGGGCGCGTCAACCCAGGCTGAGCTGACTGACTATGAAATCAGGTGTCT	222475	Db	221454	CTTGTGAATAGCTCAGAAATTGTGATCAGAAACCTGGCTGAGAAAGAAACCCCAAGAA	221335
Qy	781	TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr	800	Qy	1102	CTGTGATAGCTCAGAAATTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221275
Db	222474	TACACCTGGCCATCAACCGCTCGGGAACAATGAGCATCCCCCAAGAGTCGACACACTAC	222415	Db	221334	TCCTGGCAGCTGGCGCAGGCCCGCGCTGTGCCCCAAGGCTGGGCTGTGCTTTGGTTG	221215
Qy	801	LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgAspSerTyr	820	Qy	1102	GCCTGCCCCCTGTACAGACTGTGAGTGTGGGGGAAAGGCTTGTGAGGCTGCCCTGGC	221215
Db	222414	CTGAGAAATCTGTGTAAATGACGTGAATGACAAATGCCCTCCAGATTCCTACAGAAATCTTAC	222355	Db	221274	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155
Qy	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840	Qy	1102	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155
Db	222354	CAGGGAGTGTCTATGAGATGTGCCACTTCAACAGCCTCTGTGAGATCTACGCAAC	222295	Db	221154	GAGCTGGAGAGAGAGAGGAGGAGCTGTGGAGAGAGAGAAAGGAGGGGAGGCGGTGGGA	221095
Qy	841	AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly	860	Qy	1102	GAGCTGGAGAGAGAGAGGAGGAGCTGTGGAGAGAGAGAAAGGAGGGGAGGCGGTGGGA	221095
Db	222294	GATCGGCACTCCGCTCGAAACGGCAGGGTTTCTACACCTTCCAAAGAGGGGAGTGA	222235	Db	1102	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155
Qy	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgGluLeuAsp	880	Qy	1102	GAGCTGGAGAGAGAGAGGAGGAGCTGTGGAGAGAGAGAAAGGAGGGGAGGCGGTGGGA	221095
Db	222234	GATCGGCACTTTATTTGATGATCAATCAAGGCAATGTGTGCACTGTGCGAGGCTGTGAT	222175	Db	1102	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155
Qy	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspTyrGlyMetProPro	900	Qy	1102	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155
Db	222174	CGTGAAGATGTGGCCCAAGTACATCTTGAGAGCGTACGCGGTGACAAAGGGGAGTGGCGCA	222115	Db	221094	AGACCTTGCTGGCAGGGGATGGGTGTGAGGGGAGAGCAATCGTTCCTCAGACCTGTG	221035
				Qy	1102	TCCTGCTCCAAGAATTTGTGAAAAAAGGGCGCTGTGAAAGCCTTGAAAGCCCA	221155

Db	221034	TCACCCCTCATGGCACCCTGTGTATCTCA	CACAGCCTTTGGCAGGCTCAGAGGCTAC	CAGAG	220975
Qy	1102	-----	-----	-----	1102
Db	220974	ACACAGATTCACCAAGGGGGGTAGAGACC	CTGCGCTCCTCTCCCCCACTGTGTTTC		220915
Qy	1102	-----	-----	-----	1102
Db	220914	TCAGTAGAGCTGCAGAAACTGACTTGTG	ACTTGTCTAGACATCTTTTCCCTCCAGGC		220855
Qy	1102	-----	-----	-----	1102
Db	220854	AGAAAGAGCTGCTGGCTCCCTTCAGAGCT	CCAGCAGAGAAAGTAGAGTCAAGAGCTC		220795
Qy	1102	-----	-----	-----	1102
Db	220794	ATTGTCTGTGTCAGACAGAGCTCACTG	CCGAGATTCTTGGGTCTGGCTTGGGG		220735
Qy	1102	-----	-----	-----	1102
Db	220734	GAAAGGCTGCACAAGAGCCAGGAGATGA	ACTTTTGGGCTCCCCACCCCGTGAGCT		220675
Qy	1102	-----	-----	-----	1102
Db	220674	TGTTCTTGGGGTGTAGGTCTTCTGCTGA	CCATGATGCGATGCGCTCAGAGAAACAG		220615
Qy	1102	-----	-----	-----	1102
Db	220614	CACGGGGTTCAGAGTGCATAGAGCTAG	AGCTTTGAGGGGAGAGACTCCAGGGCCCTG		220555
Qy	1102	-----	-----	-----	1102
Db	220554	ACTGTGTGAGAGAGACAGATTCTAGGG	CGGTGAGAGCTGCGCAGAAACCAGCTGG		220495
Qy	1102	-----	-----	-----	1102
Db	220494	GCATTCTGCAGCTTGTTCAGACGCTCA	GCCTCCAGACAGAGAGCTGGGGAGGAG		220435
Qy	1102	-----	-----	-----	1102
Db	220434	CAGAGCTGGTGTATGTATCTATCCCTT	CTGTAGCTCTGAGATGAGAAAGCAAGCGG		220375
Qy	1102	-----	-----	-----	1102
Db	220374	GAGGGGAGACTCAGTAGCCAGACCTTTC	CTGAGTCTGGAGTTGGGCTGAGAGAGG		220315
Qy	1102	-----	-----	-----	1102
Db	220314	AACTGTTAAGTTTTCCAGGGGTGAGACT	ATGTGCTTATCCAGGCAGTCTTACATC		220255
Qy	1102	-----	-----	-----	1102
Db	220254	AGGGAGGGAAAGATGAGCTTGCGCCCTT	TCTGCTCATACTTCTTACTGACAGACATT		220195
Qy	1102	-----	-----	-----	1102
Db	220194	CCTGATCCACCCCTCAGGTGCTGCAGAA	AGACAGAGAGAGCTCTCCAGGCTATG		220135
Qy	1102	-----	-----	-----	1102
Db	220134	GCTTGTCAAGCTGTAAACCACTTGAAG	CTCAGTTCCCTCCAGAGAAAGAGACTAG		220075
Qy	1102	-----	-----	-----	1102
Db	220074	GACCCCTCAGCTGAGCTTGACAGAGAG	CAGAGCGCTCTCCCTCCGCTCTCCCA		220015
Qy	1102	-----	-----	-----	1102
Db	220014	TCACAGAGCTTCAGTGTGAATCCGGCT	AGATTCTTAGGGCCGTGTGAGAGACAGCT		219955
Qy	1102	-----	-----	-----	1102

---

Db	219954	GCTTACCTTTGTCTTCTCTCACAGGGG	CACAGCCCACTGATCAAACTCTAGATAG		218995
Qy	1102	-----	-----	-----	1102
Db	218954	GGAAAGGCTTGGCCGACGTAGGGCCACT	GTGAGCTGCTTTGTGCGCTTGTCCCTCC		218835
Qy	1102	-----	-----	-----	1102
Db	218834	CCATCAGCCCTGACCTTAGAGTCACTCA	CCAGGCAACCCAAAGTTCAAGAGGCTGG		218775
Qy	1102	-----	-----	-----	1102
Db	218774	TGAGAGGACTGAGGCCCATGCTCCTTGT	CTTCTCTGTGTTCTTGTCTCTCTCT		218715
Qy	1102	-----	-----	-----	1102
Db	218714	TCCCTCCACCCCTGACCTTAGGGAGCC	CTCTGCACTGCTGTCTGTCACTTCTCCCA		218655
Qy	1102	-----	-----	-----	1102
Db	218654	CTTCTGTGCTTTCCTTCTATGCTTCTC	CACTGTGCTTGTGCTTTCCTGTTC		218595
Qy	1102	-----	-----	-----	1102
Db	218594	TCTCTATGTTAGATCCTGTCTTCTCTC	TCTCACTGTGCTGAGCTGCCAGCCTGG		218535
Qy	1102	-----	-----	-----	1102
Db	218534	CTCCATGGCTCTGAGCCCAATGTCGCC	GGCTGTATCTCTTCAAGACACAGAGAA		218475
Qy	1102	-----	-----	-----	1102
Db	218474	CTTGCACAGGCTGTCACTCCAGACAG	AGTCTCTATCAAGAAAGCAAGGTTG		218415
Qy	1102	-----	-----	-----	1102
Db	218414	GAGGCTGACTGTGGCGCTTTCCTGTG	AGCTGTGAGGCTTAGCCAGGTATAT		218355
Qy	1102	-----	-----	-----	1102
Db	218354	TGCTGGGTGAGAGCCTTACCCAGCTCT	CTCTTCTTAAACAAGTCAGGCCACTCT		218295
Qy	1102	-----	-----	-----	1102
Db	218294	CCAACTGCTTGTGGAAGAAAGCAATG	AGAGAGAGAGACGATGGTGTGCTG		218235
Qy	1102	-----	-----	-----	1102
Db	218234	GAATGGGACATGACAGAGGGCAAGTGC	ACAGATGTGGCCAGATGGGAGAGAGC		218175
Qy	1102	-----	-----	-----	1102
Db	218174	AAAGATATGCCGGGCAAGGGGACAGTC	AGGCTGTGGCTGTGAACCATGGGGAGTG		218115
Qy	1102	-----	-----	-----	1102
Db	218114	TGGCTTCCAGGACAGAGATGCAAAAG	ATTCTTGTCTGTGCTTCAAGAGACTGT		218055
Qy	1102	-----	-----	-----	1102
Db	218054	TTGTAGATCCCTGTGTTCTTCTGTGA	GGGAAACATCCATCCAGAAAACATGGGC		218995
Qy	1102	-----	-----	-----	1102
Db	218994	TGTATCTTCAAGCTCTGAGCCAAATG	CAAGGCTTCTGACAGAGGCAAGAGTCTG		218935
Qy	1102	-----	-----	-----	1102
Db	218934	ACCATGTGACCCATGAGGAGCTTGAG	GGGCTTAGAGGGCTGTGACTGTGTGTA		218875
Qy	1102	-----	-----	-----	1102
Db	218874	CCCTGGCAAGAGAGAGAGAGAGTGA	CGACGTCATGCGACAGGCTTGTGAGAA		218815

[illegible]

QY	1102	-----	1102
Db	217734	CCGCCACACACCGGTGAGAGCGGGCTTATCTGTGGGTTGAGGTGACGGGCGAGATC	217675
QY	1102	-----	1102
Db	217674	TCTCTGTCTCTGGGCCCTCTAGAAAAGAAAGAAAGACAAAGCGTTTGGAGCAG	217615
QY	1102	-----	1102
Db	217614	GAAATCCAGGCGATCGCAGGGGACAGTATATCTGAATAGGGTGAAAGCAGAACT	217555
QY	1102	-----	1102
Db	217554	GCAAGCAATTAGTCAAGCTTCAATGCCAAGAGTGTGGGAAAAACAAAGACGTTT	217495
QY	1102	-----	1102
Db	217494	CGCAATGCTCTTCTTCTTAAAGCATATAGAGTAGCCCTCAAGAGAGACTGTCCA	217435
QY	1102	-----	1102
Db	217434	CAAGTGCACACAGCTCATTTGGCCACTTATAGCCATCTAGAAAAAGTTTCTTGGTACC	217375
QY	1102	-----	1102
Db	217374	TTGCCGATATAGGCTAGATTATTTAGCTGTTCACCTTCTATTCGGATATCTCTT	217315
QY	1102	-----	1102
Db	217314	ATGTGCTCCTCTTGTGCTCTCTCTCTCGGCTCCTCTTCTTAGACTCTTCTTC	217255
QY	1102	-----	1102
Db	217254	CTGTGCTCTCTCCTCGGGGCTCCTCTTGTATCTCGTTTCTTACTGCTCTTACT	217195
QY	1102	-----	1102
Db	217194	CCTGATGCTCTCTCCTCGGTGCTCCTCTTCTCAATCTCTCCTGTGTCTCC	217135
QY	1102	-----	1102
Db	217134	TCTGACACACTCCTCCTCTCTGCTCGTTTCTTACTCTCTCTGACAGCTTAC	217075
QY	1102	-----	1102
Db	217074	TCCCTCTTCCTTAAGCCAGGCTTGGCGCATCTGCTTGCCTACCCCAAGCTGA	217015
QY	1103	-----SerAPGlyValHisSerValThrAlaGlnCysAlaLe	1115
Db	217014	TGTTGATTCGTGTGCTTCTCTTCAAGATGATCTCACAGTGAAGCCAGAGCTTACT	216955
QY	1115	uArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuAs	1135
Db	216954	CGGTGTCACTATCATCACAGAGATGCTCACACAGCATCAAGCTGCGGTGGAAAG	216895
QY	1135	pMetSerProGluArgPheLeuSerProLeuGluGlyLeuPheIleGlnAlaValAlaI	1155
Db	216894	CATGTCTTCAGAAAGCTTTCTGTCCCACTGTGGACCTTTCATTCAAGCTGTGGCAC	216835
QY	1155	aThrLeuAlaThrProProAspHisValValPheAsnValGlnArgAspThrAspAl	1175
Db	216834	CACATTTGGCCACACCCCAAGTACAGTGTGTCTTCAATGTGCAGAGGAATATCTATGC	216775
QY	1175	aProGlyIleHisIleAsnValSerLeuSerValGlyGlnProProGlyIleProGlyI	1195
Db	216774	CCCAAGGCGGCATATCTCAACGTAGGCTGTCAAGTGGCCAGCCCCCAGACGCCGGGG	216715
QY	1195	YGIYProProPheLeuProSerGluAspLeuGlnGluArgLeuTYrLeuAsnArgSerLe	1215
Db	216714	TGGGCAACCCCTTTCACCTTCAAGAGATCTGCAGAGGCCCTGTATCTTCAACCGACCT	216655
QY	1215	uLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsnIleCysIleuArgI	1235



D	216654	GCTCACCCGCACCTCAGCTCAGCCGGTCTCTCCCTTCGACGACAACTTTGGCTGGAGAA	216595
Q	1235	UProCyseGluuantiYmeChagCyseValSerValIleuAtrPheAseSerAlaProPh	1255
D	216594	ACCCTGCAGAAATTACATGACCGCTGTGTCTCCGTGTGCGCTTCACACTCTCTGGCCCTT	216535
Q	1255	eIlleAseSerSerValIleuPheAtrProIleHtIProValGlvYleuAtrCyseAtr	1275
D	216554	TATGCTCTCTCTTCTGTGCTCTCTTCCGCCCACTCACCTGTCCGGGGGTCTGGCGTGTCC	216475
Q	1275	gCyseProProGlvYpHeThrglvYasPlyYCyseGlvuThrglvValAsePheuCyElyrSeAtr	1295
D	216474	CTGCCCAACAGGCTTCACAGCGGACTACTGCGAGACTGAGGTGAGCCTCTGTTAACTCAAG	216415
Q	1295	gProCyseGlvYpHeThrglvYasPlyYCyseGlvuThrglvValAsePheuCyElyrSeAtr	1315
D	216414	ACCTGTGTGAGACCCCAATGGGCACTGTCCGACGTGAGAGGGGTGTATACGTGCTGTGTCCG	216355
Q	1315	gAspGlvYlyThr-----	1319
D	216354	CGATGGCTACACGGGTGTAGTCTAGGGGACAGGACAAATGGGGGACGCTGCAGGGCTGACG	216295
Q	1319	-----	1319
D	216294	CTGTGGCCCAAGCACAAATCAGACTAATATATGGGGCTGTCTTACTTCAAGGGTATA	216235
Q	1319	-----	1319
D	216234	AACTAGATGACACAGCCGCTCTTATAGATCTACTTGTGTGTACAACTGCTGCTTATATA	216175
Q	1319	-----	1319
D	216174	AGTCAGATTAAGAACTTATGTCTACTGTGACAGCCACACACCACACTGGCAGATGG	216115
Q	1319	-----	1319
D	216114	CGCCTGGCAGACCCCAAGAACTGTGTATATGAGTCAATCCGACACTGGGTTTGT	216055
Q	1319	-----	1319
D	216054	GTTCCAGCTTACCACTAATTAATCTGTGAATAATGACGAACTACTTAGCTTCTTCCAC	215995
Q	1319	-----	1319
D	215994	AGTATTTAAGATGAGATTAACATTTATGTATGTAGGTCAAAAGTTCTGTGAGACATCATG	215935
Q	1319	-----	1319
D	215934	TAAATTAATTTGATGTAGAAAGAAAGCTGGCTTATGTGTGATGTGTGGCTTTGTTTAG	215875
Q	1319	-----	1319
D	215874	GAAGTCTGAGATGAGGTGCTTGATCCAGTTAAGAGGTGTCTGGCCTAAGAACAGGAC	215815
Q	1319	-----	1319
D	215814	CTTCATTAATGCTCTGTGAATCCTTCATTAATTAATTTAGCCAGACTTCCCAACAGAGAA	215755
Q	1319	-----	1319
D	215754	CTGAGAGAGGCACTGAGAGAGCAGGGGTTTGTAGTCAATGGGAATTCAGAGACACTGGACC	215695
Q	1319	-----	1319
D	215694	CGTGTAGCCCTGCAAAATGGCTCAATTTCTAATCCCAATTTGTCTGGGGCTGACAGTCT	215635
Q	1319	-----	1319
D	215634	CGTTACACCTGCCCAACCTGTATAGGCTGTGTCTCTCTCAGCTCTTGTCTGTCTTCC	215575
Q	1319	-----	1319
D	215574	TTCCCAAGACTCCCAAGGCTTACAGGTCTTCAGCCGCAAGGCGCACTGGCTGTGCCAGTCCG	215515
Q	1319	-----	1319
D	215514	TGGGACAGCTCAGTGCCTCTGTGTGACACCTCTGTCTCAGAACTTGCACAGACTTACTCT	215455
Q	1319	-----	1319
D	215454	TCTGTGTCTGTGTGTGTGAGGGTGGGGGTGGGGCTAGGGGAGGGGAGTCTTGTCTC	215395
Q	1319	-----	1319
D	215394	AGAGAGATGGGGAAGGGGGCCTCCGAGGTATAGGAGTAACAGTTTCCATGAAACAAGA	215335
Q	1319	-----	1319
D	215334	GCCAGTCACTAGGACAGGGAACAGAACTGAAATCTATCTTAGAATCAGAAATCGAGGAGAA	215275
Q	1319	-----	1319
D	215274	AGGAGTGGATTTTCCCGTCTAATATCTGAATTTTACAGCCCTGATTTCTGTCTGTCTC	215215
Q	1319	-----	1319
D	215214	TCAGGCTTAGGCCCCCAGTCTCTCAGCCTGTTTATGGCGGAGGGATCTCAGTTCAAG	215155
Q	1319	-----	1319
D	215154	CCCCAATGTTCTGTTCTTGGGGAAGCCGTGCTTACTTACACATTCAGACCCCAACTAT	215095
Q	1319	-----	1319
D	215094	AACTCAATATGACCTGGAAACCTATAGTCAATCTCTTAGGGAAGAAACTGATCTCTAGCC	215035
Q	1319	-----	1319
D	215034	TGACAGGAACAAAGACAGAGTGGGAAGGGGGGAAGAACGGAGTGGGGCCGGCGCTG	214975
Q	1319	-----	1319
D	214974	GAACCTGTGAGGGGCAAGGCTTAGGAGCAAGGGTGACTGACAGGTGTGGGAAGGGCCC	214915
Q	1319	-----	1319
D	214914	AAGAGCTCACAGCAGAACTTTGTGTGAGCTGCAGAGAGATAGTGTGCGACCAATTAAG	214855
Q	1319	-----	1319
D	214854	AGGAGTGTGTGCTGATTAATTTTCAAGATGTCAACAGAGGAAGTGAGAACGATGGGA	214795
Q	1319	-----	1319
D	214794	GCCCGAGATGAGAGAGGTCAACAGAGATGTTAACAGGACACGTTAGACTTCTCACT	214735
Q	1319	-----	1319
D	214734	ACTGATTAACCAAGAGAGACAGTTTAAACAAGAAAGTTTACTTGTCTCCGGTCTTC	214675
Q	1319	-----	1319
D	214674	TGCAGTTTCAAGTTCTGTGTGCTGTGTCTTCCCTATTAAGACAGACATCAGGGTGGCGAAG	214615
Q	1319	-----	1319
D	214614	CTGTCAACCAATGAGAAAGAACCTTGAGCAAAAGTATGACCTGTATGGCTACTCCCA	214555
Q	1319	-----	1319
D	214554	GTAAGTTGCTTTCAGCTCTCTCTCTCCAGATGCTTCCAGTGCCTAATAACAGCCCCC	214495
Q	1319	-----	1319
D	214494	TCAGCCAGACCAAGCCTTATGATATATGAGCTGGGGGAAGGGGGCAATTCATATTCAAAC	214435

[illegible]

QY	1462	-----	1462
Db	213354	CAGCGCTGGAGAAAGAGGAAAGATGTGCTTTCAGAGGCACTTCACATCCCTGCTGTGCTCT	213295
QY	1462	-----	1462
Db	213294	TGATGTGGGAATATGTGAAGGAGGATGTGTCCTGCTGGGTGTGCTACTGCTTCTTT	213235
QY	1463	-----ProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlyGlnGly	1478
Db	213234	CTGTGTCCACAGCCGCTGTGTGGGTGACAGGACTTCCACAGGCCCATCTGACGAAAG	213175
QY	1479	ValAlaValAlaThrValAlaSpGlyCyaaPThrGlyValAlaLeuArPheGlySerVal	1498
Db	213174	GTGGGTGTGTGTCCGTGGATGCGCTGTACACAGGGGTGCGCTGTGCGCTTCGAGGTATG	213115
QY	1499	LeuGlyAenTySerCyaaAlaAlaGlnGlyThrGlnGlyGlySerTyLeys	1515
Db	213114	CTGGCAACTACTCTGTGTGCTGCCAGGGCACCCCAAGGAGGACGACAAAGATTAAGCAGAG	213055
QY	1515	-----	1515
Db	213054	ATGGGGCGGGAGGGGGAGGGGTGCAGAGCTTAGGAGAGGGTACTGTGGCTCAACAGG	212955
QY	1515	-----	1515
Db	212994	ATGGGGCTTCAGAGTCAACGAGATGGGGCTCCGAGTAGGACTACTCAGATGACCCCTT	212935
QY	1515	-----	1515
Db	212934	CCCTACAAATGTTGGGAATTAGTCAGGTCTTGTGTGGCTGTGAGGGAGAGGAGACACATC	212875
QY	1515	-----	1515
Db	212874	TGTCAGTACTGGGGTGGCGCTGTGCTTCCTGGGGTTTTCTAAAGGGCGTCATGGCTCG	212815
QY	1515	-----	1515
Db	212814	GCTCAAGTCTTCCTTTCTTGTGCTTAGTGCAAGGTATGGGTAAAGAAACACAG	212755
QY	1516	-----SerLeuArPheThr	1520
Db	212754	ACTTAAATTCCTGTCCGAGTCTTCTTCACTCTTCCTGCCCCCAGAGTCTCTGACCTGAC	212695
QY	1520	tGlyProLeuLeuLeuGlyGlyValProArPheProGlySerPheProValArgMetAr	1540
Db	212694	AGGGCCCTTGTGTGGGGTGGGTGCCAGATCTGCCCGAGAGTTTCCCTGTCCGAATGCG	212635
QY	1540	gGlnPheValGlyCyMetArGAsnLeuGlnValArSerArgGlnIleArPheCAlaAr	1560
Db	212634	GCACTTTGTGGGCTCATGAAACCTCCAGGTGTATACCCGACGTCGACATGACCCGA	212575
QY	1560	pPheIleAlaAsnAnglyThrValProGly	1570
Db	212574	CTTCATCCGCAACATATGCACTGTGCTCG--TATAGACACTGGGGGTACGGGCAAGGCT	212516
QY	1570	-----	1570
Db	212515	GGAGACCAATATGCATGAGATCGGTTGTGCATGACACGGGAATGCCAGCTGCTGTGTT	212456
QY	1570	-----	1570
Db	212455	CTTCCCGTCACTGCTGTGTGTTTCCAGAGCCCTGGGGTCCCTGAGTCCCTCTCTCCGT	212396
QY	1571	-----CyProAlaTyValyAsnValCyaaPserAenThrCyaaIAsnGlyGlyTh	1588
Db	212395	TCTAGGCTGCCCCCAAGAAAGACGTGTGTACAGCAACACTTGCATTAATGTGTGCAC	212336
QY	1588	tCyValAnglnrPArPAlaPheSerCyMetTyCyProLeuGlyPheGlyGlyTySer	1608
Db	212335	CTGTGTGAACCAGTGGAGCGCAATTCAGCTGCCAGATGTCCCTTAGGCTTGGGGGCAAGAG	212276
QY	1608	tCyaaIAsn	1611

```

Db      212275 CTGTGCCCCAGGGTAGAGATGGGACCTGGCCACAGGTTGGGGCCTGGGCTCTGTCACTTGTTC 212216
Qy      1611 ----- 1611
Db      212215 TGGGATTACGGGAGAGGCTTGGGGCAGGCTCTGGGAGGGGCTCCGGAGTCAAGGCTGTGG 212156
Qy      1611 ----- 1611
Db      212155 GTGGAGAGCGTGTCTGGGACAGTAAAGAGAGCTAGGACAGAGACAAAGACCTAGTC 212096
Qy      1611 ----- 1611
Db      212095 AAGAGCAGGCGCTGGGACAAATGTAGGGCGGCGCTGGAAACAGGTAAATGCTCCGGTGGGC 212036
Qy      1611 ----- 1611
Db      212035 GGGGAGTAGCAGAGACCGGGTGTAGAGGGAGGAGGGTCCATCTCTCATGGAGCTGCCA 211976
Qy      1612 ----- 1614
Db      211975 CGCTTCTGATCTCACTGCTCCCTGCTGATCACTTTTCATCTGCTCCAGAAATGG 211916
Qy      1614 1AAsnProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProI 1634
Db      211915 CCATATCCCAAGCGCTTCTTGTGGGACAGCGCTGTGGCTGGGCAATGAGCTCTCTCTGCCCCA 211856
Qy      1634 1eSerGlnProTrpTyrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuL 1654
Db      211855 TCTCTCAGGCGCTGGGACCGTCAAGCTCATATGTTCCGGACACGCCAGGCAAGATGGCGTCTGTC 211796
Qy      1654 euGlnAlaAlaLeuThrArgGlyArgSerThrIleThrLeu----- 1666
Db      211795 TGCAGGCGGTCAACAGAGGGGCGCAGGACCATCACCTCGAGGTGATGACAGGCGGCTGGG 211736
Qy      1666 ----- 1666
Db      211735 AGTGTGGCAAGCCCGGCGTGGCTAACTTTGTAGAGACCTTCAGAGGGGTTGGAACCCAGGT 211676
Qy      1666 ----- 1666
Db      211675 GAGCTGGGACAGGTTTGGCAGAGACCAAGCATTTGGCTGGGCCCATTCGGGTGGTCACTACT 211616
Qy      1666 ----- 1666
Db      211615 GCCACCTATTGGGCTTAAAGAAATTAATCCCTCTCTGCTGACGCGGACAGACTGG 211556
Qy      1667 ----- 1671
Db      211555 AAGAGCTACTGAGGTGCTTAATTTTCTTTCTCTCTCTGTAACGCTTCGGGACG 211496
Qy      1671 1HisValMetLeuSerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnP 1691
Db      211495 GCCACGTAGTGTCTAATGTGTGAGGGGACAGGGCTCAGGGCTCATCCCTGGGTCTGGAGC 211436
Qy      1691 roGlyArgAlaAsnAspGlyAspTrpHisHisAlaGlnLeuAlaLeuGlyValAspGlyG 1711
Db      211435 CAGGCGGAGCCATATATGTGTGACTGGCATACGACAGCTGTCACTGGGAGGCTACCGGGG 211376
Qy      1711 1PProGlyHisAlaIleLeuSerPheAspTyrGlyGlnGlnArgAlaGluGlyAsnLeuG 1731
Db      211375 GCCCTGGCCACCGCCATCTCTCTCTTGAATGAGGACAGAGGAGGAGGTAATCTGG 211316
Qy      1731 1PProArgLeuHisGlyLeuHisIleLeuSerAsnIleThrValGlyGlyIleProGlyProA 1751
Db      211315 GCCCTCGGCTGACCGGGCTGACCTGAGCAATATATGACGTGGGGAGATTCCCGGGCAG 211256
Qy      1751 1aGlyGlyValAlaAlaArgGlyPheArgGlyCys----- 1761
Db      211255 CCAGAGTGTGGCCCGTGGCTTCGGGGCTGTGTTCAGGTGAATCTCTCCCGACCTT 211196
Qy      1762 ----- 1763

```

```

Db      211195 CCATCCATTCCTCACTGGAGCCCTGTTCCACTGAGGGCTAAATGCTCTTAATCCCTTC 211136
Qy      1763 1nglyValArgValSerAspThrProGluGlyValAlaSerLeuAspProSerHisGlyG 1783
Db      211135 AGGGTGTAGGGTAAAGCAAGACACCCAGAGGTTTAAAGAGTGGATCCAGCCGTGGGG 211076
Qy      1783 1uSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsnProCysP 1803
Db      211075 AAGACATCAATGTGAGGACGAGGCTGTAGCTGGCCAGATCCCTGTGACTCGAATCCATGTC 211016
Qy      1803 roAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp----- 1820
Db      211015 CCACCAAGCTACTGACAGCAACACTGGAGCACTATTTCTTGTAGCTGTATCCAGGTA 210956
Qy      1820 ----- 1820
Db      210955 AGCTGAGGATGCTGGAAATGGGGCTGTGTGCACGGGTCAACATCCGCTCTAATATGTA 210896
Qy      1820 ----- 1820
Db      210895 GCCAGTTACATCCAGGGTATAGCCATCTGTGTCAGGGTTCAGACGGAGTAATAAGGA 210836
Qy      1820 ----- 1820
Db      210835 GTTGTAGATGAAAGAAATATCCAGTAAGGGCGGGTCTGAAACCTTTCGATCTTCTT 210776
Qy      1821 --ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnHisG 1840
Db      210775 GTCCAGGTTACTAATGTGTGACACCTGTACAAATGTGTGTGACTGTACACCATGCGAGCAC 210716
Qy      1840 1uSerValCysThrArgGlyAspProSerAlaProHisGlyTyrThrCysGluCysProProA 1860
Db      210715 AGCTGTGTGTATCCGAAACCCAGTGCACCCGCTCACTCATCTGAGGTGTATCCAA 210656
Qy      1860 snTyrLeuGlyProTyrCysGluThrArgIle----- 1870
Db      210655 ATTACTTTGGCCCTATATGTAGACACAGTA--GGTGGCGGGGGCTTAACCTCTGTATCAC 210597
Qy      1870 ----- 1870
Db      210596 TCAGCCCCGAGATCGAGATCGGGCTGAGCTGACGCCAATGAAGAATAATATGTGTT 210537
Qy      1870 ----- 1870
Db      210536 GAACCAAGCGGAGAGATGCAGTGTAGAGGGGCTCATGCTGTGGTGTCTTCT 210477
Qy      1871 -----AspGlnProCysProAspGlyTyrTrpGlyHisIleProThrCysGlyProCysA 1888
Db      210476 GCAGGATTAACCAACCTTGCCCCGCTGGCTGTGGGAGACACCCCAATGTGTCCATGCA 210417
Qy      1888 snCysAspValSerTyrGlyPheAspProAspCysAsnIleThrSerGlyGlyCysHisG 1908
Db      210416 ACTCGAATGTACAGAAAGGCTTTGACCAGACTGCAACAAAGACAAAGTGGGAGTGCACCT 210357
Qy      1908 ybIys----- 1909
Db      210356 GCAAGGTGAGAGACCCGAAATGAGCCCCCGTGTGTACCTGTCTCTAGCTTCGCCCC 210297
Qy      1909 ----- 1909
Db      210296 TACAGGGGAAAGAGAGATGGGCTTCTCTCAAGAGCGAGACAGCAATCCTGTGCTT 210237
Qy      1909 ----- 1909
Db      210236 TTGCCTTATCAGGATGAGAGGTGTGTTTGCCCTTATATCCGGCTGACAGGGGATCCA 210177
Qy      1909 ----- 1909
Db      210176 GTTCCAGTGTAGTCAGTGAGCTGGAATACTAATGTGTGGCTGTGTCAGCCCATTCG 210117
Qy      1909 ----- 1909
Db      210116 AGTATTTGGCTTCTCACTGCTATGTCAATTACCTGCTGTGTAGAGTGTGAATCTCTCCC 210057

```

Qy	1909	-----	1909	-----	Qy	2004	-----	2004	-----
Db	210056	CACAAAGACCAGCATAGGGAGGAGCTCTCTGTGGGAGGCCAAGCTTGAAGCTCTGCT	209997	-----	Db	208976	CTCTGACACTGGGCTTCCTGTCTCCCTAAGCCCGCCTTCTTGCTTCTTCTCA	208917	-----
Qy	1910	-----	1913	-----	Qy	2005	GIYThrAlaValArgHisCysAspGluHisArgGlyIlePheProAsnLeuPheAsn	2024	-----
Db	209996	AGCATCCAGAAAGACATCTTGATCCAGTGAACCCCTTCTCCCTCAAGAGAAATCACT	209937	-----	Db	208916	GGGACTGTGTGGCCACTGTGATGAGACAGGGGGCTGGCTCCGCCAAACCTCTCAAC	208857	-----
Qy	1913	YrArgProGlySerProThrCysLeuLeuCysAspCysIleProThrGlySerLeus	1933	-----	Qy	2025	CysThrSerIleThrPheSerGlyIleuLeuGlyPheAla	2037	-----
Db	209936	ACCGACCCCCAGAGGCCCACTGGCTCTGTGTGATCTGTAAACCCCACTGGCTCTTGT	209877	-----	Db	208856	TGCAGTCAGTCACTTCTCAGAGCTAAAGGCTTTGT-AGTTGACCCCTGTCTCTC	208798	-----
Qy	1933	eTrArgValCysAspProGluAspGlyGlnCysProCysIysProGlyValIleGlyArg	1953	-----	Qy	2037	-----	2037	-----
Db	209876	CCCGAGCTGTGACCTCGAGAGATGGCCAGTGTCCAGCAAGCCTGAGTCATTTGGCGCTC	209817	-----	Db	208797	CTTCTTTCATCCCTCGTCTTTGAAAAATAGAGTCCCGGGGTTGGGGAATTAGCTAGT	208738	-----
Qy	1953	IncYsaAspArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysGlu	1970	-----	Qy	2037	-----	2037	-----
Db	209816	AGTGTATCTCTGTGACACCCCTTTGTGTAGGTCAACACATAGCTGTGAGGTAGGG	209757	-----	Db	208737	GGTAAGCGCTTGCTTAGAAGAGCGAGGCGCTGGTTCGGTCCAGCTCCGGGAAAAA	208678	-----
Qy	1970	-----	1970	-----	Qy	2037	-----	2037	-----
Db	209756	GGGCTTTAGACAGAGTAGGTCCCTCTCACCATAATGCCAAGAGCTTACCTCAGAG	209697	-----	Db	208677	AGAACCAAAAAAAAAAAAAAAAAAATAGAGCCGGAATCTGTTTCTGAGCTCT	208618	-----
Qy	1970	-----	1970	-----	Qy	2037	-----	2037	-----
Db	209696	TTAGAAATAGGCTGCATTTGGGTGGGAGTGTGAAAGAGAGTTGCCGGGATGATAG	209637	-----	Db	208617	CCTACCTGACAGGCCACTTCTTCCCACTGATCTTTGAGGCCCTTTACTGTGAGCC	208558	-----
Qy	1971	-----	1979	-----	Qy	2038	-----	2039	-----
Db	209636	AACCTGTTACCGATTAACCATGTGTCTTCTCCATAGTGAATTAATGACAGCTGCCAAGG	209577	-----	Db	208557	CCAGTCCCAACATCTGTGTGACCCCTGTGACATTCGTGTCTTCATCCAGCCGAGCGG	208498	-----
Qy	1979	IaIleGluAlaGlyIleTPTTPProArgThrArgPheGlyLeuProAlaIalaIaProc	1999	-----	Qy	2040	LeuGlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnIleuAlaLeuLeu	2059	-----
Db	209576	CCATAGAGGCTGGGATCTGGTGGCCCCGACGGGTTGGGCTACCTGCTGTGCCCCCT	209517	-----	Db	208497	CTACAGCGAAAGATTCAGGGCTGAGCTCAGAGAGCTCCAGAGGCTGGCCCTGCTCTG	208438	-----
Qy	1999	YsProIlysgIyserPhe	2004	-----	Qy	2060	ArgAsnAlaThrGlnHisThrAlaGlyIleThrPheGlySerAspValIysValAlaTyGln	2079	-----
Db	209516	GCCCCAAGGGCTCTCTTGTGTAGGTTTAAATTAACAATGATGATTAAGAACCTTGA	209457	-----	Db	208437	CGTAATGCCACACACACACACCTCTGGCTACTTCGCACTGCACTGATGTCAGAGTGGCTTACAG	208378	-----
Qy	2004	-----	2004	-----	Qy	2080	LeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThr	2099	-----
Db	209456	ACTCTCAGAGAAATAAGATAGCAGCTAGAGCCAGCTTGAAGCTCCCGGGGCCCACTT	209397	-----	Db	208377	CTGGCCACACGACTGCTGGCTCATGAGAGTCCACAGCGGGCTTTGGGCTGTCCGCCACA	208318	-----
Qy	2004	-----	2004	-----	Qy	2100	GlnAspValHisPheThrGlu	2106	-----
Db	209396	TCTGGAATTCATCTCCAGAGACTCAGTGTCTCACTTACCTTCACTCACTCAACCATCA	209337	-----	Db	208317	CAGATGTGCACCTTCACTGAGGTGGGCTTGAAGTACATGTGGGGCCAGTGAGAGA	208258	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209336	CTTAGGCTGACAAAGTGCCTTATTCAGACCACACAGAGCTGACGAAATCTGCGCTG	209277	-----	Db	208257	GCCTGTCTGTGAGATCTTTGGGGAGATCTTGAAGGAGATCTTCTACTGTAAAGAGC	208198	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209276	TCTCGAACAGGGCTGGAGCTAGACTGTGTGATTCATGGGGTGCTATCTTGAAGCTAAC	209217	-----	Db	208197	TGCCCAACCCCTGTACGAGGCTTCATGATGATCTGTAAACAGAGCGACAACAAGTTG	208138	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209216	TACTTAGTACTGAGATTACCCCACTGAAAGGAGGCCGTGCTGAGAGTATGGCGTCT	209157	-----	Db	208137	ATTAAGATACCTCTCTGTGGGCTGGGGAATTAGTCACTGATGAGCGCTTACTTAG	208078	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209156	GGGTACTATCCCAAGATTCACATTAGTGTGACANGTTCATGGGCAAGGAGTACT	209097	-----	Db	208077	AAGCGAAGGCCCTGGGTTGGGTCCCAAGCTCCGAAAAAAGAAAAAAGAGTA	208018	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209096	GAGAGTAAGCAGTAGCTGAGTGAAGAGCGGTAGTATCTTTGGGGCGGTGTGAGAGA	209037	-----	Db	208017	CTTCTCTTACCAAGTTCAGGAATTAGTACTTGCATTTGAGGGCTGACCTGAGGCA	207958	-----
Qy	2004	-----	2004	-----	Qy	2106	-----	2106	-----
Db	209036	GGTTACATAGCGTCCCTTTGGCCACTGTACTGATTTTGTCTGAGAGTGTTTGGCTGTGAG	208977	-----	Db	207957	GAGCCCACTATTATACCAAGAGCTCAGTCTTGTCAAGGAGCCAAAGAGGCAAAAAACA	207898	-----
Qy	209036	GGTTACATAGCGTCCCTTTGGCCACTGTACTGATTTTGTCTGAGAGTGTTTGGCTGTGAG	208977	-----	Qy	2106	-----	2106	-----

Db 207897 GAGCGAGACATGCAAGGACACATGGGTGGCCCTGATTAAAGTGTGAAGAGATTAGA 207838  
 QY 2106 ----- 2106  
 Db 207837 GGGCGAGTACCATGCTGGAAGTGAACAGTATGATACAAAGGCTTCCAGGGGAACTG 207778  
 QY 2106 ----- 2106  
 Db 207777 TGGGACTGCTAAGGAGAGACCTAGTGGCGGGTCTGAACAGGACAGGCTGAACCTT 207718  
 QY 2107 ----- 2107  
 Db 207717 GGTATAGGTTACCAACAGCATATGCCACTGTGACCAAGCCCATCTCCAGATCTGCTG 207658  
 QY 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisTrpGluLeuIleGlnGln 2129  
 Db 207657 AGGTGGGAGAGCGCCCTCTCGTGAATGGCGCAATGAAGGCACTGGGAATGATCCAGCAG 207598  
 QY 2130 ThrGluGlyGlyThrAlaTrpLeuLeuGlnHisTrpGluValLysSerAlaLeuAla 2149  
 Db 207597 ACAGAGGGTGGCACTGCTGCTGGCTGCTCCAGCACTATGAGGCTTATGCCAGTGGCCCTGGCC 207538  
 QY 2150 GlnAsnMetArgHisThrTrpLeuSerProPheThrIleValThrProAsnIle----- 2167  
 Db 207537 CAGATATGCGGCAATCACTAAGAGCCCTTCACTATGCTCACACCCAAACATTGGTGA 207478  
 QY 2167 ----- 2167  
 Db 207477 GTGGTCTGGGCTGGGGGCGGGGTGTGGAAGTCTGCCCATTTCAAAGGCGAGCCAG 207418  
 QY 2168 ----- 2168  
 Db 207417 CCGTGTCTGGTGAAGTCACTTGCACCAATATTTCTTCTCTAGTCACTCTGTGGGCGCCTG 207358  
 QY 2175 AspLysGlyAsnPheAlaGlyAlaLysLeuProArgTrpGluAlaLeuArgGlyGluGln 2194  
 Db 207357 GATTAAGGGAACTTGTCTGAGCAACAGCTGCCCGTATGAGGCGGTAGCGTGGAGCGCC 207298  
 QY 2195 ProProAspLeuGluThrThrValIleLeuProGluSerValPheArg----- 2210  
 Db 207297 CCTCCAGATCTTGAGACCAAGTCACTTTTGGCCAGAGTCTGTCTTCAGAGTCAAGCGGGGA 207238  
 QY 2210 ----- 2210  
 Db 207237 AGGTGATGTGTAAAGCGCTGGGGTGGGGCAACCATGCAAGTGGGCGAGCGCAGTGA 207178  
 QY 2211 ----- 2211  
 Db 207177 GTAGCAGCATTTACTTGCAGAGATGCTCCCATGTGAGATCTGCAGGACCTGTGTGAAC 207118  
 QY 2224 aglnGluProGluGluLeuAlaArgArgLysArgArgHisProGluLeuSerGlnGlyGly 2244  
 Db 207117 CCAGGAGACTGAGAGAGCTGGCCCGGGGAGAGGAGGCAACCGAACTAGATCAGGGGCA 207058  
 QY 2244 uAlaValAlaSerValIleIleTrpArgThrLeuAlaGlyLeuLeuProHisAsnTrpAs 2264  
 Db 207057 GGCAGTGGCGAGTGTCACTATTTACCAACGCTGGCTGAGTGTGCCCAACATAGCA 206998  
 QY 2264 pProAspLysArgSerLeu----- 2270  
 Db 206997 CCGAGTAAAGCGGTAGCTGAGGTAAAGCGGCTACGACAGGCTTAGAGAGTTGGGGTTAGC 206938  
 QY 2270 ----- 2270  
 Db 206937 AAGGAATGTGAAGCATCTCCAGTGAATAGGATGCTCTGCAAGCCGCTTTGTGACTGT 206878  
 QY 2271 ----- 2271  
 Db 206877 GTACCTTCTGTCAATCAGAGTCCCAAGCGCCAGCATCAACACCTGTGTGAAGCAT 206818  
 QY 2285 eSerValHisAspArgGluGluLeuProArgAlaLeuAspLysProValThrValGly 2305

Db 206817 CAGTGTCCAGATGATGAGAGCTCTCTCCAGCGGCACTGGAACAAGCAGTCAAGTGA 206758  
 QY 2305 nPheArgLeuLeuGluThrGluGluArgThrLysProIleLysValPheTrpAsnHisSe 2325  
 Db 206757 GTTCCGACTGTGGAGAGGAGAGGAGCAACAGCCCATCTGTGTCTTGTGAACCATTC 206698  
 QY 2325 rIleLeu----- 2327  
 Db 206697 GATCCGTGAGCCTGCTGTACCACTTAAGCCCATGACTTTGAGCAAGATGCCAGGC 206638  
 QY 2328 ----- 2328  
 Db 206637 TCCGTGCTGTCACTTTCTTCTCTCCAGTCCGCTCTTCTCTGCAAGGCTCAAGTGGCA 206578  
 QY 2331 hrcGlyGlyTrpSerAlaArgGlyCysGluValValPheArgAsnGluSerHisValSerC 2351  
 Db 206577 CAGTGGCTGTGCTCCCGAGAGCTCGAGGTTGTCTCCGTAACAGAGCCATGTCAAGCT 206518  
 QY 2351 yGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArgArgGlu----- 2369  
 Db 206517 GCCAGTGCACCAATATGACAAAGCTTTGCTGTGCTCATGGAATGTATCCGAGAGAGTGG 206458  
 QY 2369 ----- 2369  
 Db 206457 GACTCCAGGGGTAGCCCAAGAGCAGGGGTGGGCAACAAATGCAAGGGCTTCAGGCCCTG 206398  
 QY 2370 ----- 2370  
 Db 206397 CTCTTTTGAACCCCAACCACTCATCTCCTCAGAAATGTGAGATTTTGGCACTGAAGACCC 206338  
 QY 2379 eutThrTrpValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeuT 2399  
 Db 206337 TGAAGTATGTGGCCCTTGGAGTACCTTAGCTGCGCTGATGATCACTTCCCTTCTCA 206278  
 QY 2399 hrlLeuLeuArgGlyLeuLeuArgSerAsnGlnHisGlyLysArgArgAsnLeuThrAlaAla 2419  
 Db 206277 CCTCTTTCAGAGCCCTTGTCTTCCCAACGATGGATGCCAGCAACTCCACAGCTGCCC 206218  
 QY 2419 eugLysLeuAlaGlnLeuValPheLeuLeuGlyLysAsnGlnAlaAspLeuPro----- 2436  
 Db 206217 TGGGCTGGCCCAAGTGTCTTCTCTGGGAGATCAACAGGCTGACCT-CCCTGTAAGA 206159  
 QY 2436 ----- 2436  
 Db 206158 TACTTCTCTCCAGAGACTTCCCAACTTCAAGCCCAAGGCTCATGTGTCCCTT 206099  
 QY 2436 ----- 2436  
 Db 206098 CTATGACCTTGTCCATCCCTGCTCCCAACAGAAATCAATAGTAAAGGCCCATTCAGT 206039  
 QY 2436 ----- 2436  
 Db 206038 AACTGCCCAAGTTGGCCCTTCTCTCAACACAGAGCTTCCCACTCCCTGGGGGAGC 205979  
 QY 2437 ----- 2437  
 Db 205978 TCCTTGTATGACTGTCTGCTGCTGCTGATCATGTGCTGCGCCCTCAAGTTTCTTGACAG 205919  
 QY 2441 alIleAlaIleLeuLeuHisPheLeuLysLeuCysThrPheSerThrAlaLeuLeuGluAla 2461  
 Db 205918 TCAATGCCATCTCTGTGACTTCTGTACCTTTGACACTTCTCTGGGCTGTGTGAGG 205859  
 QY 2461 lAlaLeuHisLeuTrpArgAlaLeuThrGluValAspArgValAsnThrGlyProMetArgP 2481  
 Db 205858 CCTTACACCTGTACCGGGCGGTTCACAGAGGTGCGAGAGCTGAAATGCCAGGCCCATGCGTT 205799  
 QY 2481 heLysTrpMetLeuGlyTrpGlyValProAlaPheIleThr----- 2494  
 Db 205798 TCTACTACATGCTGGGGTGGGGCGTCTCTGCTTCAATACAGATATCTGCTCTTCC 205739  
 QY 2494 ----- 2494  
 Db 205738 TGACCTAGAGCTCCCTGTGTGTAGGCTGGGGGTCTACCTTCAATGATGATACCCCAAC 205679

QY	2494	-----	-----	-----	2494
Db	205678	CATGAATTCCTGAGAAATCCCTTCCCTTAATCTTGGAACAGCCTGGATTAGGTGTACGGG			205619
QY	2494	-----	-----	-----	2494
Db	205618	GCTAGGCTTCTCTTGGAGAGGGGAGTGGACAGAGGCTCTTCTGTATGTCTCCCACTTG			205559
QY	2494	-----	-----	-----	2494
Db	205558	GCTCTGTGAACCTTGAATGATCATACTTGCTCCACCTTGAGACCTGTCTTGCTGCT			205499
QY	2495	-----	-GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAsp		2510
Db	205498	GTCCTGTGCTCCAGAGTCTGGCTGTGGGCTTGATCCCAAGGCTATGGAAACCTGACT			205459
QY	2510	heCyTrpLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPhe			2530
Db	205438	TCTGTGTGCTCTCCATCTACATACAGTACGCTCATCTGAGGTTTGTGTGACCAAGTGGCTTTG			205379
QY	2530	laValSer-----			2532
Db	205378	CTGTTTCGGTAAAGCTGCTGAAGCAATTAGGGTAAACAGATGGACGGGTCTACAGAT			205319
QY	2532	-----	-----	-----	2532
Db	205318	TCCCTTTAGAGAGAGCTGAGGCTCTGGGCCCTCGGGCAAGAAAGTCTCTGCTGC			205259
QY	2533	-----	-MetSerValPheLeuTyrIleLeuAlaAla		2543
Db	205258	TCTCTGTCTCTCATATGTCCCACTGACAGATGATGTCTCTCGTATATCTGTGCGCCC			205199
QY	2543	rgAlaSerCyValAlaIaGlnArgGlnGlyPheGlyuIysIysGlyProVal-----			2559
Db	205198	GAGCTCTGTGTCTGCCCCACGTCAGGGCTTCGAGAAAGAAAGGCCCTGTGTAGTACTGG			205139
QY	2559	-----	-----	-----	2559
Db	205138	CAGGTGCTGGGACAGGCGGGGTGTGGTATAGAGAGCATCATGGCGGGTGTGGTAT			205079
QY	2560	-----	-SerIleuGlnProSer		2565
Db	205078	GAGAGCATCATGGCGGAATCATCATATGCCCTGCCCTCCAGCTCAAGGCCCTGCTCTCC			205019
QY	2566	PheAlaValLeuLeuLeuLeuSerAlaThrTyrLeuLeuAlaLeuLeuSerValaAsnSer			2585
Db	205018	TTCACTGTCTCTCTGTGCTGATGTGCATATGGCTGCTGGCACTGGCTGTGTGTCAACAGC			204959
QY	2586	AspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle-----			2600
Db	204958	GACACTCTCCCTCTTCACATCACTCTTCGGGCTGTGAATTTGT-CCAAGTACCTAGCCC			204900
QY	2600	-----	-----	-----	2600
Db	204899	CGCCTTACTGAGTGGAGAGAGCATGCTCTGGGGGTGGGCTGTGCTCAACATAGCTGTCT			204844
QY	2601	-----	-GlnGlyProPheIlePheLeuSerTyrValIaLeuSerIysGluValaArglys		2618
Db	204839	GCCTTCACAGGGCCCTCTCATCTTCCCTCTCTACGTGTGTCTGACGAAGGAGGTCCGAAAA			204780
QY	2619	AlaLeuLysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrTyrLysSer			2638
Db	204779	GCACTCAAGTTGGCTCTGCAGCGCGAAGCCAGCCTCTGACCTTGTACCATTAAGTCT			204720
QY	2639	ThrLeuThrSer-----			2642
Db	204719	ACCTGACCTC-GTGTAGGGGGCGGGGTCTGAGCTCAAGAGCGCTTGTGTGAAGGGG			204661
QY	2642	-----	-----	-----	2642
Db	204660	AAGCAGGCCCCAGAGAACTGTGTGTCTTTGTATCTTTTAACTGGGCACTCTTTTCCC			204601

[illegible]

```

Db 203532 TCCCTGGGCGAGGAGGCTGGAGACGCTCTGGGCTCTGAGCTGAGACGCTGCTTC 203473
Qy 2775 1eSerThrProlys----- 2779
Db 203472 ACAGTACCCCCCAA-GGGTGGGCGACGACCAAGGGCTGACGACATTGGGGGGGCGACAGGAAG 203414
Qy 2779 ----- 2779
Db 203413 CCTAAGAGCTAGAGGCTTGTGGCTATCTCTGCTACCTCTCTGAGGACTATAGGCTTCC 203354
Qy 2780 -----AspGlyGlyProGlyProGlyLysAlaProThrProGlyAspPheGlyT 2796
Db 203353 TTCTCTTGACAGATGGGGGTCCAGGGTCTGGGAGAGTCCCTTGGCCAGGAGCTTGGGA 203294
Qy 2796 hTThrAlaLysGlySerSerGlyAsnGlyAlaProGlyLysAlaProGlyLysAlaGly 2816
Db 203293 CCACAACAAAGAGAAATAGTGTAGTGGGCGCCCTTGAAAGAGGGCGACAGGAGGAGTGGAG 203234
Qy 2816 spAlaLysSerArgGlyLysSerGlyProLeuProGlySerSerAlaGlnProHisL 2836
Db 203233 ATGCTCTTACCTGGAGAGGGTCTCTGGGAGCCCTTCCGGGGCCCTTCTACCCAACTCA 203174
Qy 2836 YSGLYLe----- 2838
Db 203173 AAGGTTA-GTGAAGAGTCTCTCTCATCTGCTTCACTTCCGATCTCTTGTGGAATTTT 203115
Qy 2839 -----LeuLysLysCysL 2844
Db 203114 AGAATGCTCCATCTCATGTGTACCAACACTCTTACCCAGGAGCTCTCAAGAGAGAGTGT 203055
Qy 2844 euProThrLysSerGlyLysSerSerLeuLeuArgLeuProLeuGlyLysCysThrGlyS 2864
Db 203054 TGCCCAACATCAAGCAAAAGATAGGCTCTCTAAGTGTGCTTGAAGAGGAGGAGGAGT 202995
Qy 2864 eSerArgGlySerSerAlaSerGlyLysArgGlyGlyProProArgProProP 2884
Db 202994 CTTCCCGGGGCTTCCACCGCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 202935
Qy 2884 rGArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2904
Db 202934 CGCGCGAGTCTCTCCAGGAGAGCGTGAATGGGTATGCGCATCCGATGAGCATCAAGG 202875
Qy 2904 laGlyThrValAspGluAspSerSerGlySerGlu 2915
Db 202874 CAGGCACTGTGATGAGGAGCTCTTGTGGCTCCGAG 202840

RESULT 14
AC095125/c AC095125 254975 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-8N10, WORKING DRAFT SEQUENCE, 2
unorderded pieces.
ACCESSION AC095125
VERSION AC095125.6 GI:30467550
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 254975)
Muzny,D,Marie, Metzger,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Altschuler, S., Amin, A., Angilano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Caret, K., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Dejgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITL
JOURNAL
COMMENT
JOURNAL

```

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gevorgian, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guera, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Huix, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, T., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovs, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzen, L., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minj, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokelimeh, O., Okwou, G., Olariunsaagou, A., Pal, S., Parks, K., Paeternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, B., Pu, L., Pu, M., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smales, J., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Stratek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, J., Wei, X., White, P., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 254975)  
Worley, K. C.  
Direct Submission  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 254975)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 9, 2003 this sequence version replaced gi:2940723.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by extended gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web Site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GCNC  
Center clone name: CH230-8N10  
----- Summary Statistics



Assembly program: Atlas;  
 Consensus quality: 231495 bases at least Q40  
 Consensus quality: 232513 bases at least Q30  
 Consensus quality: 233358 bases at least Q20  
 Estimated insert size: 240763; sum-of-coverage estimation  
 Quality coverage: 10x in Q20 bases; sum-of-coverage estimation

\*\*\*\*\*  
 NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 253729: contig of 253729 bp in length  
 \* 253730 253829: gap of unknown length  
 \* 253830 254975: contig of 1146 bp in length.  
 Location/Qualifiers  
 1. 254975

## FEATURES

source  
 /organism="Hattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-8N10"  
 misc\_feature  
 /note="1.1872  
 misc\_feature  
 /note="wgs contig"  
 misc\_feature  
 /note="wgs contig"  
 /note="wgs contig"  
 BASE COUNT 60412 a 54867 c 56326 g 62413 t 20957 others  
 ORIGIN

## Alignment Scores:

Prod. No.: 0 Length: 254975  
 Percent Similarity: 11983.00 Matches: 2789  
 Best Local Similarity: 38.64% Conservative: 47  
 Query Match: 38.00% Mismatches: 75  
 DB: 2 Indels: 4437  
 Gaps: 33

US-09-916-849A-3 (1-2923) x AC095125 (1-254975)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 242219 ATGCGGAGCGGGGCTCCAGCGGCCCTCCCAAGCCACTGCTGCTACTGCTG 242160  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 242159 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242100  
 QY 41 GlySerArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCysProSer 60  
 Db 242099 GGGTCCGGGGGAGCGCGCTCTCTGCGGCTGCGGCCCTCGGCTGCTGCGCCAGCC 242040  
 QY 61 SerAlaSerAsnLeuTrpLeuTrpThrSerArgCysArgAspAlaGlyThrGluLeuThr 80  
 Db 242039 TCTGCTTCAACCTCTGCGCTTACACCAAGCGCTGAGAGATTGCGGCACTTGAAGCTGACC 241980  
 QY 81 GlyHisLeuValProHisAspGlyLeuArgValTrpCysProGluSerGluAlaHis 100  
 Db 241979 GGCACCTGAGTCCCATCAACGAGGCTGAGGCTTGGTGTGTCACAAACAGGGGCTCAT 241920  
 QY 101 IleProLeuProProAlaProGluGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
 Db 241919 ATCCCTCTCCGCGCATCTCTGAGAGCTGCGCTGAGCTGTGACTCTCGGATACGGA 241860  
 QY 121 GlyHisLeuSerProGluGlyCysLeuThrLeuProGluGluHisProCysLeuLeuAla 140  
 Db 241859 GAGCACCTTCTTCCAAAGGCAAGCTGACCTGCTCAAGAACACCTTGTAAAGGCC 241800  
 QY 141 ProArgLeuArgCysArgSerCysLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160

Db 241799 CCACGGCTCAGATGCCAGTCTTGCAGCTGCGACAGGCCGCCAGGGCTGCGGGA 241740  
 QY 161 ArgSerProGluGluSerLeuGlyGlyArgArgValAsnValAsnThrAlaProGln 180  
 Db 241739 GGATACACAGAGAAATCTATGGGTGGGCGAGAAAGAAATGAAATACAGCTCCCAAG 241680  
 QY 181 PheGlnProProSerTrpGlnAlaThrValProGluAsnGlnProAlaGlyThrProVal 200  
 Db 241679 TTCCAGCTCCCAAGTACCAAGCCAGTCCGAGAACCAAGCCTGCTGTAACCTCTGT 241620  
 QY 201 AlaSerLeuAlaGlnIleAspProAspGluGlyGluAlaGlyArgLeuGlyTrpMet 220  
 Db 241619 GCTTCCTTAAAGGCATTGAATCCGAGTGAAGGGGTGAGGCTGCACTTGAATCAACCAATG 241560  
 QY 221 AsnAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db 241559 GATGCCCTCTTGAATAGCGGCTCCAAATCATTTCTTCTTGAACCCCAATCACTGGTCT 241500  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThr 260  
 Db 241499 GTACCAACAGCTGAGAGCTGATCGGGAGCAAGAGCAACCAATGCTTCAGGGTCAAC 241440  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db 241439 GCACAGATCATGGTATGCCCGGACGAGTCCCTTGGCCACACTTATCTTATGTACT 241380  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGluTrpLysGluSerLeuArgGlu 300  
 Db 241379 GACACCAATGATCAAGACCTGTTTGTAGCAGCAAGAAATCAAGAGAGCTTCAGGAG 241320  
 QY 301 AsnLeuGlnValGlyTrpGluValLeuThrValAlaArgAlaThrAspGlyAspAlaProPro 320  
 Db 241319 AACCTGAGGTTGGCTATGAGGTCTTACAGTCAAGGCGCACGATGGGAGGCCCTCT 241260  
 QY 321 AsnAlaAsnIleLeuTrpArgLeuLeuGlnGlySerGlySerProSerGluValPhe 340  
 Db 241259 AATGCACATCTGTACCGCTCTCTGAGGGGCTTGAAGGACCCCTCAAGAGTCTT 241200  
 QY 341 GluIleAspProAspSerGlyValIleArgThrArgGlyProValAspArgGluGluVal 360  
 Db 241199 GAGATCATCTCTGCTGCGGCTGATCCGACCCGCGGCTGATGATCGGAGGAAGTGT 241140  
 QY 361 GluSerTrpGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380  
 Db 241139 GAATCTTCAAAATGATGAGTGGAGCGAGTACCAAGGCTCGGAGCCCAAGGCGGAGGC 241080  
 QY 381 ThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400  
 Db 241079 TCCAGCTGTCGCTCTCTGCTGCTGAGATGATGATGACAGCCGCCCAAGTTAGC 241020  
 QY 401 GluIleAspArgValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420  
 Db 241019 GAGAAAGGCTACGATGAGTTCAGGTGCGGAGGAGATGACCCCAAGAGCCCGGTTCTCGA 240960  
 QY 421 ValThrAlaSerAspAspAspGlyGlySerAsnAlaValHisTrpSerIleMetSer 440  
 Db 240959 GTCAAGGCTTGGCAAGAGCAAGGCGGCAAGCCCTGCGGCTATGATGATGATGAGC 240900  
 QY 441 GlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460  
 Db 240899 GGCACGCTCGGGGACAGTTTATCTGATCTGACACTGAGAGCCCTGATGTGGTAGC 240840  
 QY 461 ProLeuAspTrpGluThrThrLysGluTrpThrLeuArgValArgAlaGlnAspGlyGly 480  
 Db 240839 CCACTCGACTATGAGACCAACAAAGAAATATCACTGGGAGATCCCGGCGCAGAGATGGTGC 240780  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db 240779 CGGCTTCACCTTCCAAAGCTCCGATCTAGTAACCGGAGGCTCTAGACATCAATGAC 240720  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 240719 AAGGCCCACTTGTGTACAGACCCCTTCCAGGCGCAAGGCTCTTAGAGATGATCTTTA 240660

OY	521	GIYTYrLeuValleuHiValGlnAlaIAspAlaAspAlaGlyAspAsnAlaArgLeu	540
Db	240659	GGCTCACTGGTCTCGCATGTCGAAGCGATTGATGTCGATGCGGAGGAAACGCCGCTTC	2406000
OY	541	GIYTYrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTYrGly	560
Db	240599	GAATCAAGCTCTGCTGGGAGTGGGACAGACTTCCCTTCACTTAACAACGACACAGGC	2405400
OY	561	TPPIleSerValAlaIleGluLeuAspArgGluGluValAspPheTYrSerPheGlyVal	580
Db	240539	TGGATCTCGTGTGGCCCGGAGTTGGACCGGGAAAGAGTTGATTCTACAGCTTCGAGTA	2404800
OY	581	GIuAlaArgAspHisGlyTYrProAlaLeuThrAlaSerAlaSerValSerValIThrVal	600
Db	240479	GAAGCGAGGACCAACGCGACCCCAAGACCTCACCGGCTCAGCGTCAAGTGAACATC	2404200
OY	601	LeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTYrThrValArgLeuAsn	620
Db	240419	CTGGATGTCATGAACAACAACCCACCTTTACGACCGGAGTACACGGGCGGCTCAAC	2403600
OY	621	GIuAspAlaAlaValGlyTYrSerValIThrValSerAlaValAspArgAspAlaHis	640
Db	240359	GAGGACGGCGTGTGGGACCAAGTGTGGTGAAGGTGTCAAGCGGTGATGAGATGCTCAC	2403000
OY	641	SerValIleThrTYrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660
Db	240299	AGTGCATTATTCCTACCAATCAACCAAGGCAACACCGGCAACCGCTTCTCTACACAGC	2402400
OY	661	GIuSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTYrIleuGluArgGln	680
Db	240239	CAAAACGGGGGTGGCTGTCTCCCTGGCTTACCGCTGGATCAAACTCGAGCGGCAAG	2401800
OY	681	TYrValLeuAlaValIThrAlaSerAspGlyTYrArgGluAspThrAlaGlnIleValI	700
Db	240179	TATGCTGGCGGTGACTGCTCGATGGACAAGGCAAGGACCAAGCTCAAGTCTGGTGG	2401200
OY	701	AsnValIThrAspAlaAsnThrHisArgProValPheGlnSerSerHisIleTYrThrValAsn	720
Db	240119	AATGTCACGTGAAGCCAAACCCATGCTCCGCTTCCAAAGCTCCCACTAATACGGTGAAT	2400600
OY	721	ValAsnGluAspArgProAlaGlyTYrThrValIleuIleSerAlaThrAspGluAsp	740
Db	240059	GTTATATGAAGACCGGCGACGACGACCAACAAGTGTCTGATCAGTCTCAACAGCAGGAC	2400000
OY	741	ThrGlyGluAsnAlaArgIleThrTYrPheMetGluAspSerIleProGlnPheArgIle	760
Db	239999	ACAGGGGAGAAAGCCCGAATCACTACTTATAGGAGACAGCATCCCTCAAGTCCGCAATC	2399400
OY	761	AspAlaAspThrGlyAlaValIleThrArgGlnAlaGluLeuAspTYrGluAspGlyValIleSer	780
Db	239939	GATGCAGACCTGGGGCGGTCAACACCGGCTGAGCTGACATATAGATCAAGAGTGTCT	2398800
OY	781	TYrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleuSerAspThrTYr	800
Db	239879	TACACCTCGGCATCAACCGCTCGGGACAAATGGCATCCCAAGAAATGGACCAACTAC	2398200
OY	801	LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTYr	820
Db	239819	CTGGAGATCTCGGTAAATGACGTGAATGCAATGCCCCCAAGTCTCTCAGAGATTCCTAC	2397600
OY	821	GIuGlySerValTYrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	239759	CAAGGGAAGTGTATGAGGATGTGCCACCTTCAACAAGGTCTCTGAGATCTCAAGCAAC	2397000
OY	841	AspArgAspSerGlyLeuAsnGlyArgValPheTYrThrPheGlnGlyGlyAspAspGly	860
Db	239699	GATCCGCACTCCGGTCTGAACCGCAGGGGTTCCTACCTTCCAAAGAGGAGCGACATGGA	2396400
OY	861	AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrIleuArgArgLeuAsp	880
Db	239639	GATGGTGACTTATATGTGAAGTCCACATCAAGCATGTGTGGACACTGTGGAGGCTGGAT	2395800

QY	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIysGlyMetProPro	900
Dp	239579	CCTGAAGAATGTGCCCAAGTACACTCTTGAGAGCGTACCGCCTGGACAAGGGAAATCCGCCCA	2395220
QY	901	AlaArgThrPrometGluValThrValThrValLeuAspValAsnaAsnProProVal	920
Dp	239519	GCCCCCACACCATTGAAAGTGCAGTTACTGTACTGAAATGTGAATGACAAATCCCCCGTC	2394600
QY	921	PheGluGlnAspGluPheAspValPheValGluGlnSerProLleGlyLeuAlaVal	940
Dp	239459	TTTGAACAGATGAGTTGATGATATTGTGGAAAGAACAGCCCCCATCGACTGGCGTGT	2394000
QY	941	AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnLleMetTyrGlnLle	960
Dp	239399	GCCCCGATCACAGCCACCGACCACAGCGAAGGACCAATATGCACAGTCAATGTATCCAGATA	2393400
QY	961	ValGluGlnAsnLleProGluValPheGlnLeuAspLlePheSerGlyGluLeuThrAla	980
Dp	239339	GTTGAGGGGCAAAATATCTCGAAGGTCTTTCAAGCTGAGCAATTTCTCGGGCCAAGCTTAAGGGCC	2392800
QY	981	LeuValLaAspLeuAspTyrGluAspArgProGluTyrValLeuValLleGlnAlaThrSer	1000
Dp	239279	CTCGTAGATTGGACTTACAGAGAACCGACTGAATATCTCTGGTCAATCCAGGCTTACTCT	2392200
QY	1001	AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAsnPro	1020
Dp	239219	GCTCCCTTGGTGAAGAGGGGCTACTCTCATGTCCCGCTCTCGACCGCAATGATAACCG	2391600
QY	1021	ProValLleuGlnAsnPheGlnLleLeuPheAsnAsnTyrValThrAsnArgSerSerSer	1040
Dp	239159	CGAGTCTGGGCAACTTGAAGATCTTTTTCACAACTAATGTCAACCAACGCTCGACACAG	2391000
QY	1041	PheProGlyGlyAlaLleGlyArgValProAlaHisAspProAspLleSerAspSerLeu	1060
Dp	239099	TTCCTGGGGGGGTATAGGCGCGGTGCTGCCATGACCCGATATCTCAGACAGCCTG	2390400
QY	1061	ThrTyrSerPheGlnLleArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly	1080
Dp	239039	ACGTACACCTTTGACCGAGAAATTAATCACTCAACCTGGTCTCAATCCCTCCACACCGGT	2389800
QY	1081	GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaLleMetSerVal	1100
Dp	238979	GAGCTGAACATGAGCGCGGCACTGSAACAACACCGCCCTTGAAACCATCATGAGTGTG	2389200
QY	1101	LeuVal-----	1102
Dp	238919	CTGGTGTCAAGTAAAGAAAGATGCCAGTGAATGTTGGGGTGGAAATAGTCTTTAAGGAGG	2388600
QY	1102	-----	1102
Dp	238859	TCTTGAAGCACTGAGAGGTGGTGGCCTTCCAGAAGGGGCTGTGACCCAGCTTCCAC	2388000
QY	1102	-----	1102
Dp	238799	CCTGTGATAGCTCAAGAGTTGTAGATCAAGAACTGTGGCTGAGAAGAAACCCCAAGAA	2387400
QY	1102	-----	1102
Dp	238739	TCTGGCAGCTGGCGCAGGCCCCCGCTGCTGCCACAGGCTGGGCTGTGCTTTGGTTG	2386800
QY	1102	-----	1102
Dp	238679	GCTGCCCCCTGCTTACAGACTGTGACGTGGTGGGGAAAGCGTTGTAGAGCTGCCCCTGSC	2386200
QY	1102	-----	1102
Dp	238619	TCTGCTGCCCAAGATTGTGAAAAAAAAGGGCGCGTGAAGCCTTGAAGGCCCA	2385600
QY	1102	-----	1102
Dp	238559	GAGCTGGAGAGAGAGAGGGAGCTGGGAGAGAGAGAAAGGGAGGGGAGGGGCTGGGA	2385000
QY	1102	-----	1102

Db 238499 AGACCTTGCTGCGACGGGAGTGGGTGGAGGGGAGGCAATGCTTCTCAGACCCCTG 238440  
QY 1102 ----- 1102  
Db 238439 TCACCTCATGGAACCTGTGATCTCAGACAGCCTTTGGGAGGTCAAGGGCTTACAGAG 238380  
QY 1102 ----- 1102  
Db 238379 ACAAGGATTCACACAAAGGGGGTCAAGGACCTGCGCTCTCTCTCCCACTCTGGTTTC 238320  
QY 1102 ----- 1102  
Db 238319 TCAGTAGAGCTGAGAAACTGACTTGACCTTGTGTAGAGCATTCCTTTCCCTCCAGGC 238260  
QY 1102 ----- 1102  
Db 238259 AGAAGAGCCTGCTTGCTGCTCCCTTCAGAGCTCCAGGAGGAGTAGAGTCAAGCTTC 238200  
QY 1102 ----- 1102  
Db 238199 ATTGTCTGTCCAGACAGACTCACTGCGCTGAGTTCTTGTGGGGTCTGGCTTGGGG 238140  
QY 1102 ----- 1102  
Db 238139 GAAGGCTGCAACAAAGCCAGGGCAGATGGACTTTTGGGCTCCCAACCCCGTGGCT 238080  
QY 1102 ----- 1102  
Db 238079 TGTTCCTGGGCTGAGTCTTCTCTGTGACCATGATGAGCATGGCTCAGGAGAACAGA 238020  
QY 1102 ----- 1102  
Db 238019 CAGGGGTTCAAGTGCATAGGACTACTGCTCTTGAAGGAGAGATCCAGGGCCCTG 237960  
QY 1102 ----- 1102  
Db 237959 ACTTGTGAGAGAGAACAGATTTCTAGGGGCGTGGAGCTGGCCAGAACCCAGCCTGG 237900  
QY 1102 ----- 1102  
Db 237899 GCATTCCTGAGCTTGTCAAGGCTCAGCCCCCAGACAGAGAGCTGGGAGGGGAG 237840  
QY 1102 ----- 1102  
Db 237839 CAGGACTGGGTGATGATTTCTATCCCTCTGTCAAGCTCTGAGATGAGAGAGAGCGG 237780  
QY 1102 ----- 1102  
Db 237779 GAGGGGCGAGCTCAGTAGCCAAAGCCTTCTGTAGTCTGGAGTTGGGGCTGGAGAGG 237720  
QY 1102 ----- 1102  
Db 237719 AACTGTTAAGTTTTCCAGGGGGTGAAGCTATGTGGCTACTCAGGCAATCTTACATATC 237660  
QY 1102 ----- 1102  
Db 237659 AGGAGGGAAGATGCTTCGCCCTTCTGTCTATCTTACTGTAGACAGACATT 237600  
QY 1102 ----- 1102  
Db 237599 CTTGATCCACCCCTCAGTGGCTGCAAGAACAGGAGAGAGCTTCCAAAGCTATG 237540  
QY 1102 ----- 1102  
Db 237539 GCTTGTCAAGGTGTAAACACTGAAGCTCAATTCCTCCAGAGAGAGAGAGTAAAG 237480  
QY 1102 ----- 1102  
Db 237479 GCACCTCAGCTGAGCTGCGACGAGAGGAGAGAGGCTCTCCCTCCGCTTCCCA 237420  
QY 1102 ----- 1102

Db 237419 TCACAGCCTTCAGTGTGAATCCGGCTAGATTCCTAGGGCCCGTGTGAGAGACAGCTCT 237360  
QY 1102 ----- 1102  
Db 237359 GCTTAGCCTTGTCTTCTCACCAGGGCAGACAGCCCAAGTCAACCTTAGATAGAG 237300  
QY 1102 ----- 1102  
Db 237299 GGAAGGCTGGCCGACAGCTAGGGCCACTGTGGCTGCTTCTTGCCCTTGTCCCTCCC 237240  
QY 1102 ----- 1102  
Db 237239 CCATAGCCTTGACCCCTTAGTACGTACCAAGGCAACCCAAAGTTCAAGAGGCTGGG 237180  
QY 1102 ----- 1102  
Db 237179 TGAGAGGACTGAGCCCACTGTCCCTTGTCTCTCTCTCTGTCTTGTCTCTCTCT 237120  
QY 1102 ----- 1102  
Db 237119 TCCCTCCACCCCTGCCATGGGACCTCTGCACTGCTGTCTGTCACTTCTCCAG 237060  
QY 1102 ----- 1102  
Db 237059 CTTCTGTGCTTCTTCTTATGCTTCTTCACTGCTGGCTCTTGTCTTCTCTCTCT 237000  
QY 1102 ----- 1102  
Db 236999 TCTCTATGTTAGATCCGTCTTCTCTCTCTCTCTCTGCTGAGCTGCCAGCTGGG 236940  
QY 1102 ----- 1102  
Db 236939 CTCCATGCTCTCGAGGCCAAATGTCGCGCTGTATCTCTTCAACACAGAGAA 236880  
QY 1102 ----- 1102  
Db 236879 CCTGCCAAGGCTGCTCACTCCACAGAGAGTCTTATCAAGAAAGCAAGTTG 236820  
QY 1102 ----- 1102  
Db 236819 GGAGGCTGACTGTGCGTCTTCCCTGTGAGCTGGCTGAGGCTTAAACCCAGGTGATAT 236760  
QY 1102 ----- 1102  
Db 236759 TGCCTGGTCAAGGCTTACCCAGCTCTCTTCTTAAACAGTCAAGCCACTCTATTC 236700  
QY 1102 ----- 1102  
Db 236699 CCACCTGCTTCGGAAGAGAGCAATGAGAGAGAGAGCAATGGGTTGGTTCG 236640  
QY 1102 ----- 1102  
Db 236639 GGATGGGAGCAGTAGCAGAGGGGAGGTGGCCAGATGTGGGCGAGATGGGAGAGC 236580  
QY 1102 ----- 1102  
Db 236579 AAAGCATGCGGGAGAGGGGACAGTACGCTGGGCTGTGAACACCATGCGGAGTG 236520  
QY 1102 ----- 1102  
Db 236519 TGGCTTCCAGAGCAGAGAGATGGAAGATTTCTTGTCTGTGTCTTCAAGAGCAGTGT 236460  
QY 1102 ----- 1102  
Db 236459 TTGTAGAGTCCCTGTGTCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 236400  
QY 1102 ----- 1102  
Db 236399 TGTAACTTCAAGCTCTGTAGCCCAAGTCAAGGCTTCTGACAGAGGCAAGAGTCTGATC 236340  
QY 1102 ----- 1102  
Db 236339 ACCATGTGACCCATGGGAGGCTGGAGGGCTTAGAGGGCTGTGACTTGTGTCTTGGTA 236280

QY 1102 ----- 1102  
Db 236279 CCTGGGCAAGAGAGAGAGAGTGAGCAGCATGTGAGGAGAGGCTTGTGAGAA 236220  
QY 1102 ----- 1102  
Db 236219 CAAGCAGATCACACAGAGAGGTTGAAAGGCGAGTAGGCGGAGAGCCAGGACCACTG 236160  
QY 1102 ----- 1102  
Db 236159 AGATGGGTACTGCTCTTCTGATATGCCAAGCACTTGAAGTGTACATATACAGATAGG 236100  
QY 1102 ----- 1102  
Db 236039 GTCCCATATCGCTTTTCTACTTCTGCTCAGAGACAGCTGCTCAAGTCCCATCCACTA 236040  
QY 1102 ----- 1102  
Db 236039 GATGTCAATTTTGTCTTGATTGAGAGAGTTGCTCTCTCTCTCTCTCTCTCTCTC 235980  
QY 1102 ----- 1102  
Db 235979 TC 235920  
QY 1102 ----- 1102  
Db 235919 AGATGAATTTTTCGTGTGACCTCTGTCTGTCTAGAACTTGTCTGTATGCAAGGCTG 235860  
QY 1102 ----- 1102  
Db 235859 CCTTGAACTCAGAGATCTCTGCTCTGCTCTGAGCACTGTATTAAGGTAGCGGT 235800  
QY 1102 ----- 1102  
Db 235799 CACCAACAAGTTCGCTGCGACAGTTTAAAGTCTTTATATGAAGGATTTCTCATC 235740  
QY 1102 ----- 1102  
Db 235739 AGTGTGTAGTTTAACTAGGAATACCCCTATACACCTTGAAGCGTTCTCTCTCTCT 235680  
QY 1102 ----- 1102  
Db 235679 TTGCTCAGTTTCAATTTTAACTTGAGTCATTGACCTGTGTATGACATTCAAGGTGT 235620  
QY 1102 ----- 1102  
Db 235619 GCATGTACTTCTGAGCCTAGGTGCGGTGGGTGGAGTTGGGACCGTCAATGTGTACT 235560  
QY 1102 ----- 1102  
Db 235559 GTTGTGAGTGTGCGCTGCTGAGGGGTATGGCAGGAGACCTTTGGGTACACTTGGT 235500  
QY 1102 ----- 1102  
Db 235499 GAGCAAGTGAATGATGAAATTCCTAAGAGCCCTAGTGCAGAGAGAGAGCAAGCCAGT 235440  
QY 1102 ----- 1102  
Db 235439 GGGGCTCAGGGTGAAGAGCTGCCGTCCTTCTTATTTCTACCAAGATCCACCCCAACC 235380  
QY 1102 ----- 1102  
Db 235379 CCCACACCTTTCTCAGCTTGGGCACTGGCGGTGCTGACACTGGCCGAGTGTCTC 235320  
QY 1102 ----- 1102  
Db 235319 AGCTGAATTTCTTGGGTGGGAGCCCAAGCTGCCAGAGTTGATGAAGAAGGCTCGG 235260  
QY 1102 ----- 1102  
Db 235259 TGTAAAGGTGTGTGCATGGGGCTAGCAGGTTTGACGGAACAGATTATGAGCTGAGTG 235200

QY 1102 ----- 1102  
Db 235199 CAAGGGAAACAGAGTCAGTGAAGACATGTTTCCAGAACCAATTCTGTCCACACCAAC 235140  
QY 1102 ----- 1102  
Db 235139 CCGCACAGCAGGTGAGAGGGGGTTATCTGGGGTTTGTAGGTGTACGGGGCAGAGTCT 235080  
QY 1102 ----- 1102  
Db 235079 TCTCTGTCTCTGTGGGCTCTTGAAGAAAGAGCAAGAACCAAGCCGTTTGGAGCAG 235020  
QY 1102 ----- 1102  
Db 235019 GAAATCCAGGAGTCCGAGGGGACAAGTATTAAGTGAATAGGTGAAGAGCAAAACACT 234960  
QY 1102 ----- 1102  
Db 234959 GCAGGAGTTAGTTCAAGCTTCATGCCAAGAGTGTGGGAAACAACAAAGAACGTTT 234900  
QY 1102 ----- 1102  
Db 234899 CGCAATGCTCTTTTCTTAAAGCATATGACGTAGCCCTCAAGGAGAGACTTGTCCA 234840  
QY 1102 ----- 1102  
Db 234839 CAAGTACACAGCTCATTTGCCACCTTTATGCACTAGAAAAAGTTTCTTGGGTACC 234780  
QY 1102 ----- 1102  
Db 234779 TTGCCGATATGGGCTAGATTATTAAGCTGTGACCTTCTTATTCGAGATCTCTTT 234720  
QY 1102 ----- 1102  
Db 234719 ATGGATGCTCCCTTTGGTC 234660  
QY 1102 ----- 1102  
Db 234659 CTGGTGTC 234600  
QY 1102 ----- 1102  
Db 234599 CCTGATGCTC 234540  
QY 1102 ----- 1102  
Db 234539 TCTGCAGACATCCCTC 234480  
QY 1102 ----- 1102  
Db 234479 TCCCTC 234420  
QY 1103 ----- 1103  
Db 234419 TGTGATTCGTGTC 234360  
QY 1115 UARGValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAs 1135  
Db 234359 CCGTGTACCATATCACAGACGAGATGTCTACACACAGATCACAGCTGCGCTTGGAAAG 234300  
QY 1135 pMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAl 1155  
Db 234299 CATGTCTCCAGAAAGCTTTGTCTCCCACTGCTGGACTCTTCAATCAGGCTGTGGCGGC 234240  
QY 1155 aThrLeuAlaThrProPheAspHisValValAlaPheAsnValGlnArgAspThrAspAl 1175  
Db 234239 CACATTTGGCCACACCCCAAGATCACGTGGTGTGTTTCAATGTGCAGAGGATACATGATGC 234180  
QY 1175 aProGlyIleHisIleLeuAsnValSerLeuSerValGlyIleAsnProGlyProGlyIle 1195  
Db 234179 CCGAGGGGGCCATATCTCTCAACGTAAGCTGTGAGTGGGCAAGCCCAAGGACCCGGGGGG 234120  
QY 1195 yGlyProPheProSerGluAspLeuGlnIleArgLeuThrLeuAsnArgSerIle 1215

Db	234119	TTGGGCAACCCCTTTCTA	CTTCAGAGGATCTGAGAGCGCCCTGTACTCAACCGCAGCCT	234060
Qy	1215	uLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsnIleCysIleuArgSl		1235
Db	234059	GCTCAACGCCACNTCA	GCTCAACGGGTGCTCCCTTCGACGACAACTTTGCTCGGAGAA	234000
Qy	1235	uProCysArgIuAsnIlyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPh		1255
Db	233999	ACCCGCAAGATTATCA	TGCGCTGTGTGTCCGTGCTGCCTTCGACTCCTCTGCGCCCTT	233940
Qy	1255	eIleAlaSerSerSerValIleuPheArgProIleHisProValGlyIleuArgCysAr		1275
Db	233939	TATGCGTTCCTCTT	TGGTGTCTTTCCGCCCAATCCACCTGTGCGGAGCTGCGCTGTG	233880
Qy	1275	qCysProProGlyPheThrGlyAspPylrCysGluThrGlyValAspLeuCysIlySerAr		1295
Db	233879	CTGCCACCAAGGCTTCA	CAGCGGCACTACTGCGAGACTGAGGTGGAACTCTGTATTCTCAAG	233820
Qy	1285	qProCysArgIyProHisArgIyArgCysArgSerArgIuGlyIyIyTrThrCysLeuCysAr		1315
Db	233819	ACCTGTGTGAACCCCA	TGGGCACTGCGCAGTGAAGAGGTGGCTATACCTGTGCTGTGTG	233760
Qy	1315	qAspGlyTrThr		1319
Db	233759	CGATGGCTTACACGGGTGA	GTCTAAGGCAAGGACAAATGGGGGCAAGCCCTGAGGGCTGAAG	233700
Qy	1319			1319
Db	233699	CTCTGGGCAAGCAAA	TAGACTATATATGAGGCGCTGCTCATTTACTTCCAAGGATATA	233640
Qy	1319			1319
Db	233639	AACTAGATGCAACAG	CCGCTTCTTAGATCTTACTGTGTGACAACTGCGCTTAGTA	233580
Qy	1319			1319
Db	233579	AGTCAGATTAAAGAA	CTTATGTCTACTGTGCAAGCCACACACCACTGCGCATATG	233520
Qy	1319			1319
Db	233519	CGCTGCGCACAGCC	CAAGAACTGTGTATTAATGAGTCAGTCCGGCACCTGGGTTGTG	233460
Qy	1319			1319
Db	233459	GTTCCAGCTTACCA	CTAATAAATCTGTAAATAGAGCAATCTACTAGCTTCTCTCCAC	233400
Qy	1319			1319
Db	233399	AGTATTTAAGATGAG	ATACCAATTTATGTAGTCAAAAGTTGCTTGTGAGACTCATG	233340
Qy	1319			1319
Db	233339	TAAATTAATTTGAT	AGAGAAAGCTGGCTTATGTGTAGTGTGGCTTTTGTGTAG	233280
Qy	1319			1319
Db	233279	GAAGTCTGAGATG	AGTGGCTTATCCAGTTAAGAGGTGTGCTGGCCTAAGAACAGG	233220
Qy	1319			1319
Db	233219	CTTCAATATGCTCT	GTAATCCCTTCAATTAACCTTGTAGCGAAGCTTCCCAACCAAGAA	233160
Qy	1319			1319
Db	233159	CTGAGAGAAAGCA	CTGAGAGAGGGTTTGAATGACAGGAGTTCCAGGACCTGAAC	233100
Qy	1319			1319
Db	233099	CGGTGAGCCCTGCA	AAATGGCTTCATTCTACCTCCCAATTGGTCTGTGGGCTGCAAGTCT	233040
Qy	1319			1319

Db	233039	CGTTACACCCCTGCCCCACCCCTGTAGAGCTCTGCTTCTCTCAGCTCTTGTCTCTGTTTCC	2329680
QY	1319	-----	1319
Db	232979	TTCCGAGACTCCAGGCTTTCAGGTCCTCCAGGCGCAGGCGCATGGCTCTGCGAGTCCGG	2329200
QY	1319	-----	1319
Db	232919	TGGGCACTCAGTGCCTCTGTGACACCCTCTCTGCAGAACTTGCCAGACTTGACTCT	2328660
QY	1319	-----	1319
Db	232859	TCTGTGCTGCTGTTGTGTAGGGGTGGGGGTGGGGGGCTAGGGGAGGGGGAGTCTTGTTC	2328000
QY	1319	-----	1319
Db	232799	CAGAGATGGGGAAGGGGCTCCGAAGTAAAGGATTCAGTTTCATGAGAAACAAG	2327400
QY	1319	-----	1319
Db	232739	AGCCAGTCACTAGGCACAGGAACTGAGAACTATCTCTAGAACTCAGAACTCGGAGGGAG	2326600
QY	1319	-----	1319
Db	232679	AAGGAGTGAATTTCCGCTATATATCTGAATTTTCAGCCCCCTGATCTGTGTCTTC	2326200
QY	1319	-----	1319
Db	232619	CTCAGGCTTAGGCCCCCAGTCTCTCAGCCTTGTTTAGGCGGAGGAGATCTCAGTTCAG	2325600
QY	1319	-----	1319
Db	232559	AACCCAAATGTTCTGTTCTTTGGGGAAGCCCTGGCTTACTTCACACCATGACCCCAACTA	2325000
QY	1319	-----	1319
Db	232499	TAACTCAATATGACCTTGAAACTTAGTCACATCTCTAGGGAAGAAACTGATCTTGACC	2324400
QY	1319	-----	1319
Db	232439	CTGACAGGAACAAAAGCAGAGTGGGGAAGGGGGGAAGAACGGGAGTGGGGGCTCGGGCT	2323800
QY	1319	-----	1319
Db	232379	GGAATCTGACGGGCGCAGGGCTTAGGACCAAGGCTGACTGACAGTGTGGGGAAAGGCC	2323200
QY	1319	-----	1319
Db	232319	CAAGAGCTCACAGCAGAACATTTGTGAGCTGCAGAGAGAAATGTTGTGCGAGGCCAATTA	2322600
QY	1319	-----	1319
Db	232259	GAGAGTTGCTGTGCTCATTAATTTTCAGAGATGTCAACAAGAGAAGTGAAAGCATGGG	2322000
QY	1319	-----	1319
Db	232199	AGCCCGAGATGAGAGAGGTCACAGAGATGTTTACCAAGGACACGTTAGACTTCTCAC	2321400
QY	1319	-----	1319
Db	232139	TACTGGATTAACCAAGAGAGACAGTTTAAAAACAAGAAAGTTTACTTTGTCTCCGGTCT	2320800
QY	1319	-----	1319
Db	232079	CTGCAATTCAGTTCTGTGTGCTGTGCTTCCTCGCTATAAGACAGACATCAGGGTGGCGAA	2320200
QY	1319	-----	1319
Db	232019	GCTGCTCACCCAGATGAGAGAGCCCTGGGACAAAGATGAGCCCTGTATGGGTACTCCC	2319600
QY	1319	-----	1319
Db	231959	AGTGAATGCTTCTTCACAGCTCTCTCTCCAGATGCTTCAGATGCTCATTAACAAGCCCC	2319000

QY	1319	-----	-----	1319	QY	1462	-----	-----	1462		
Db	231899	CTCAGCCAGACCAAGCCTTATGATATGAGCTGGGAGAGGGGATTCATATTCAA	231840	Db	230819	ACAGAGAGCTAAGAGGATCTGTCCGTGGCTGTATTGTAGATGTGGGAGAGGCTGTTCC	230760	Db	230819	ACAGAGAGCTAAGAGGATCTGTCCGTGGCTGTATTGTAGATGTGGGAGAGGCTGTTCC	230760
QY	1319	-----	-----	1319	QY	1462	-----	-----	1462		
Db	231839	CTGTAAAGAGGAGCTTCCAGCATCATCTTCGAGGCTTCACCTGCTTCCTCC	231780	Db	230759	ACAGCGCTGGGAGAGGAGAGATGTGCTTTCAGAGGACCTTCATCCCTGTGTGCTC	230700	Db	230759	ACAGCGCTGGGAGAGGAGAGATGTGCTTTCAGAGGACCTTCATCCCTGTGTGCTC	230700
QY	1320	--GlyGluHisGlyGluValSerAlaArgSerGlyArgCysThrProGlyValCysLeuA	1339	QY	1462	-----	-----	QY	1462	-----	-----
Db	231779	CAGGGAGAGCATGTGTAAAGTGAAGTGGCCGCTCAGGCGCTGTATCTCAGAGAGCTGGA	231720	Db	230699	TTGATGTGGGAAATATGTGTAAAGCGAGATGTGCTTCGTGGTGTCTACTGCTTCCT	230640	QY	1463	-----	-----
QY	1339	ArgGlyGlyThrCysValAlaLeuLeuValGlyGlyPheLeuGlySerProSerGlyVal	1359	QY	1463	-----	-----	QY	1463	-----	-----
Db	231719	ATGGGGGTACCTGTGTCAACCTGTGTGGGAGGCTTCAAAATGTGACTGCCATCCGGGG	231660	Db	230639	TCTCTGCCACAGCCGCTGTTGGTCAACAGAACTTCACAAAGGCCCATCTGACAGCAA	230580	QY	1478	-----	-----
QY	1359	SPHegLulubProTyProTyArgGlnValThrThrArgSerPheProAlaHisSerPheLeu	1379	QY	1478	-----	-----	QY	1478	-----	-----
Db	231659	ACTTGAAGAGCCTTCTTCAGAGTGAACCAACGACGCTTCCAGCCGCTCTTCATCA	231600	Db	230579	GATGCGTGTGTGTCCGTGTATGTGCTGTGACACAGGGGTGGCTGTGGCTTCGAGACTAT	230520	QY	1498	-----	-----
QY	1379	hrPheArgGlyLeuLeuArgGlnArgPheHisPheThrLeuAlaLeuSer	1394	QY	1498	-----	-----	QY	1498	-----	-----
Db	231599	CCTTCCGGGGCTTGGCCAGCCCTTCATCTTCACTTCCAGCTGCTCTGTGTAGTGAAGGC	231540	QY	1498	-----	-----	QY	1498	-----	-----
QY	1394	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231539	CCTGGGGTTCAGAGGACCTGTGTAGAGCGTGGGGTACTCTGTGTCAAGTACGGGGTG	231480	QY	1498	-----	-----	QY	1498	-----	-----
QY	1394	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231479	GCAGATCACTGTAAGAAAGTACCAAGAGAGCTGTGACAGAGCATTTGTATGCTGTGCC	231420	QY	1498	-----	-----	QY	1498	-----	-----
QY	1394	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231419	AGCCGGCCGCGCTCACTCTCAGAGATCTGGCTGGCACTCTCCCATCTTGAAGATGGGC	231360	QY	1498	-----	-----	QY	1498	-----	-----
QY	1394	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231359	GGCCAGAAATCTTGGGTCTGTGATCTGCACTCCAGACCTTACCAAGATGTCAAGTGTG	231300	QY	1498	-----	-----	QY	1498	-----	-----
QY	1394	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231299	CTGTGCTTAAATGAGCCAGATGATGACACGCGGGGTGGTCCCGGTTCTCGGCCCGGT	231240	QY	1498	-----	-----	QY	1498	-----	-----
QY	1395	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231239	TGACCATCCCTTATACAGGTTTGTGTAACCAAGAGAGGTGACGGGCTACTGCTGTAACG	231180	QY	1498	-----	-----	QY	1498	-----	-----
QY	1408	lyArgPheArgGlnLysHisAspPheValAlaLeuGlnValIleGlnGlnValGlnL	1428	QY	1498	-----	-----	QY	1498	-----	-----
Db	231179	GGCGCTTCAATGAGAGCATGACTTGTGTGCTGTGAGGTGTATCCAGAGGACGATGACG	231120	QY	1498	-----	-----	QY	1498	-----	-----
QY	1428	euThrPheSerAla	1432	QY	1498	-----	-----	QY	1498	-----	-----
Db	231119	TCACTTCTCTCAGGTGAGCCAGCGGCTCTGTCTTGTCCCATCTCTAAAGTCTTAT	231060	QY	1498	-----	-----	QY	1498	-----	-----
QY	1432	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	231059	GTCTCCAACTCTGACCTCAAGCATGCGCTCCCTCCAGTACAGACTGTGACCTGGCC	231000	QY	1498	-----	-----	QY	1498	-----	-----
QY	1432	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	230999	TCCATCTCTAGACATGCCACTTCCCTGCACTCCACTCTCTCTGTGACCCAGCTTA	230940	QY	1498	-----	-----	QY	1498	-----	-----
QY	1433	-----	-----	QY	1498	-----	-----	QY	1498	-----	-----
Db	230939	CCGGCTCTTCTCTAGGGGAAATCCACACACGAGTGTCTCCATCTGTGCTGGAGGGGCT	230880	QY	1498	-----	-----	QY	1498	-----	-----
QY	1448	SerAspGlyGlnTrpHisThrValGlnLeuLysTyTyAsnLys	1462	QY	1498	-----	-----	QY	1498	-----	-----
Db	230879	AGTATGCGCAGTGCATACATGTAAGTGAAGTACTCAATATGATGATGGGATGGGGGGG	230820	QY	1498	-----	-----	QY	1498	-----	-----

Db	229740	CTGTGTGAACCAAGTGGGAACGATTCAGCTGGAGATGTCCCTTAGCTTCGGGGGCAAGA	229681
Qy	1608	erCyba1ag1n-----	1611
Db	229680	GCTGTGTCCAGGGGTAGAGTGGCAGCTGCACAGGTTGGGGCTTGCTGTCAATTGT	229621
Qy	1611	-----	1611
Db	229620	CTGGGATTACGGGCAAGGCGCTTGGGGCAGGCTCTGGGGCAGGCTCCGCGAGTCAGGCTGTG	229561
Qy	1611	-----	1611
Db	229560	GGTGAAGAAGCGTGTCTGGGCAAGTAAAGAGACTAAGCAGAGACAAGAAGCTTAGT	229501
Qy	1611	-----	1611
Db	229500	CAGAGCAGGCGCTTGAGGCAACAATTAGAGGCGGGGCTTGAAACAAGTAATGCTCCGGTGGG	229441
Qy	1611	-----	1611
Db	229440	CGGGGCAATAGCGAGGACCGGGTGTAGAGGGAGGCGAGGGTCCATCTCATAGAGCTGCC	229381
Qy	1612	-----GlnMet	1613
Db	229380	ACGCTTCTGACTCACTGCCCTTGCCCTTGTAATCACTTTTTCATCTGACCTTCCAGAAATG	229321
Qy	1614	AlAsnProdlrh1sPheLeug1ysSerLeuVal1aTPH1sG1yLeuSerLeuPro	1633
Db	229320	GCCAAATCCCAAGCGTTTCTTGGGACAGACCTTGTGGCTGGCATGGCTCTCTCTGCCC	229261
Qy	1634	IlSerGlnProTTPyTrLeuSerLeuMetPheArgThrArg1na1AspG1yVal1Leu	1653
Db	229260	ATCTCTACGCGCTGGGACCTCAAGCCTCAATGTTCCGCAACAAGCAGAGCAGATGGGCTCTG	229201
Qy	1654	LeuGlnAla1leThzArg1yArgSerThr1leThLeu-----	1666
Db	229200	CTGCAAGCGGTCAACAGAGGGGGCGACACAATCACTCTGCAAGTGAATGCAGGGCGGCTGG	229141
Qy	1666	-----	1666
Db	229140	GAGTGTGGCAAGCCCGGCTTGGCTTAACTTGTAGAGACCTTACAGGGGTTTGAACCAAG	229081
Qy	1666	-----	1666
Db	229080	TGGCTGGGACAGGTTTGGACAAAGCCAGACATTTGGCTGGGCCATTTGGTGGTCAAGTAC	229021
Qy	1666	-----	1666
Db	229020	TGCAACCTATTGGGCTTAAAGAAATATACCTCTCTCTGCCCCAGCGGGGACACTG	228961
Qy	1667	-----GlnLeuArgGln	1670
Db	228960	GAAAGAGCTACTGAGGTGTCTACTAATTTTTTCTTTCCTCTCTGTACACTTCGGGCA	228901
Qy	1671	G1yH1sValMetLeuSerValG1uG1yThrG1yLeuGlnAlaSerSerLeuArgLeuGln	1690
Db	228900	GGCCACCTAATGTCTTAAGTGTGAAGGGCAACAGGGCTCCAGGCTCACTCCGTCTGTGAG	228841
Qy	1691	ProG1yArg1laAsnAspG1yAspTTPH1sH1sAlaGlnLeuAlaLeuG1yAlaSerG1y	1710
Db	228840	CCAGGCGGAGCCCAATATGATGTGACTGCATCAACGCAAGCTGTCACTGGAGGCTAAGGGG	228781
Qy	1711	G1yProG1yH1sAla1leLeuSerPheAspTTPyG1yGlnGlnArgAlaG1uG1yAsnLeu	1730
Db	228780	GGCCCTGTGGCAACGCAATCTGTCTTTTGTACTATGGGCAACAAGAGGGTAAATCTG	228721
Qy	1731	G1yProArgLeuH1sG1yLeuH1sLeuSerAsn1leThrValG1yG1y1eProG1yPro	1750
Db	228720	GGCCCTGGGTGCACCGGGCTGCACCTGAACCAATATGACAGTCCGGGGAGTTCCGGGGCA	228661
Qy	1751	AlaG1yG1yVal1aArgG1yPheArgG1yCys-----	1761

Db	228660	GCACGAGGTGTGGCCCGTGGCTTCCGGGGCTGTGTTGACAGTGAATCTCCTCCCCGCACT	228661
QY	1762	-----Leu	1762
Db	228600	TCCATCCCATTCCTACCTCGAGCCCTGTGTCCACTGAGGGCTAATCTGCTTAACTCCCTT	228541
QY	1763	GLNGLYVALAAGVALSERASPThr-ProGluGlyValIAsnSerLeuaspProSerHISGly	1782
Db	228540	CAGGTGTGAGGGTGAAGCAACCAACCCAGGGTGTGATGACGTCTGGATCCAGCCGTGGG	228481
QY	1783	GLuSerIleasnValGluGluGlyCysSerLeuProaspProCysaspSerxanProCys	1802
Db	228480	GAGAGCATCATGTGAGAGCCAGCGCTGTAGCTGGGCCAGATCCCTGACCTCAATCATGT	228421
QY	1803	ProIleasnSerTyrcysSerxanaspThraspSerTyrcysCysSerCysasp-----	1820
Db	228420	CCACCAACAGCTACTGCAAGACAGACTGGGACAGCTATTCTTGTAGCTGTATCCAGGT	228361
QY	1820	-----	1820
Db	228360	AAGCTGAGGATCTGGGAAATGGGGCTGTGTGCACAGGGTCACTCCGTCCTAATATGT	228301
QY	1820	-----	1820
Db	228300	AGCCAGTTACATCCAGGGTATAGCCATCTGTGACAGGGTTACAGCGGAATAAATAGG	228241
QY	1820	-----	1820
Db	228240	AGTTGTTAGATGGAAGAGAAATGCCAGTAAAGGGGGGTCTGTAACCTTCCGATCCTTCT	228181
QY	1821	---ProGlyTyTYrGlyYaspAsnCySerThrasnValCysaspLeuasnProCysgluHis	1839
Db	228180	TGTCCAGGTACTACTGTGGAGAACACTGTGCAATGTGTGTGACTGAACCCATGCGAGCAC	228121
QY	1840	GLuSerValCysThrAspGlyProSerIleProHisGlyTYrThrCysgluCysProPro	1859
Db	228120	CAGTCTGTGTGTACCCGAAACCCAGTGCACCCACCGGTACATCTGCAAGTGTTTACCA	228061
QY	1860	AsnTyrluenglProTyrcysgluThrArgIle-----	1870
Db	228060	AATTACTTGGGGCCTATTGTGAGACAGGTA-GGTGGGGGGGGCTTACTCTGTATCA	228002
QY	1870	-----	1870
Db	228001	CTCAGCCCGGAGAGTCGAGAGTCGGGCTGAGCTGAGCCATGAAGAATACTAATGTGT	227942
QY	1870	-----	1870
Db	227941	TGAACCAAGCGGAGAGATGCAATGATGTAAGGGGGGCTCATGCTGTTGGGTGTTTC	227882
QY	1871	-----AspGluProCysProArgGlyTYrThrGlyHisProThrCysgluProCys	1887
Db	227881	TGCAGGATTGACCAACTTGTGCCCCGTGTGTGGTGGGACACCCCAAGTGTGTCATGC	227822
QY	1888	AsnCyaspValSerIySerGlyPheaspProaspCysasnLysThrSerGlyGluCysHis	1907
Db	227821	AATCGCATGTCAAGCAAGGCTTGAACCCAGACTCAACACAGACAAAGTGGCGAGTCCAC	227762
QY	1908	CysLys-----	1909
Db	227761	TGCAAGGTGAGAGACCAGATGAGCCCCGCGTGTGTGTAACCTGTCTGTAGCTTCTGACC	227702
QY	1909	-----	1909
Db	227701	CTACAGGGGAAGCAGAGGATGGGCTTCTCTCCAGAGAGCGGAGCAGACATCTGTGCT	227642
QY	1909	-----	1909
Db	227641	TTTGCTTATACAGGATGAGGAGGTGTGTTGTCCTTATATCCGGCTGACAGGGGATCC	227582
QY	1909	-----	1909
Db	227581	AGTTTCAGTGTGAGTCAAGTGAAGCTGAGCTAGCTATGTGGGCTGTTTGCAGGCCATTC	227522



QY	1909	-----	-----	-----	1909
Db	227521	GAGTATTGGCTTCTCACTGCTATGTCATTACCTGCTGTGAGAGNGTGAATCCTCTCC			227462
QY	1909	-----	-----	-----	1909
Db	227461	CCACAGAGCCGACATAGAGGAGGCTCTCTGGGAGGCCACAGCTTGAAAGCTCTGCG			227402
QY	1910	-----	-----	-----	1912
Db	227401	TAGCATCCGAGAAAGCATCTTGATCAGTGAACCCCTTCTCTCCCTCCTAGAGAAATCAC			227342
QY	1913	TyrArgProProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeu			1932
Db	227341	TACCGACCCCGCAGCAGCCCACTTGCTCTGTGTGACGTGTACCCCACTGGCTTTTG			227282
QY	1933	SerArgValCysAspProGlyLysArgGlyGlnCysProCysLysProGlyValIleGlyArg			1952
Db	227281	TCCCGAGTCTGTGACCTTAGAGATGGCCAGTGCATGCAAGCTGGAGTCAATGGGCGT			227222
QY	1953	GlnCysAspArgCysAspAsnProPheAlaGlyValThrThrAsnGlyCysGlyLeu			1970
Db	227221	CAGTGTGATCGCTGTGACACCTTTTGTGAGGTACACCAATGGCTGTGAAGGTAGG			227162
QY	1970	-----	-----	-----	1970
Db	227161	GGGGCTTTAGCAGAGTAGTTCCTCTCTCAACATATGCCAAGAGCTTACACTCCAGAG			227102
QY	1970	-----	-----	-----	1970
Db	227101	GTTAGAAATGAGCTGGCATTTGGGTGGGAGATGTGAGAGAAAGTTGGCGGCGCATGATA			227042
QY	1971	-----	-----	-----	1978
Db	227041	GAACTGCTTACCGGATACCATGTGTTCTTCTCATATGAAATTAAGACAGCTGGCCAGG			226982
QY	1979	AlaIleGlyAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaIlePro			1998
Db	226981	GCCATAGAGGCTGGGATCTGGTGCCCGCCAGCGGGTTGGCTACCTGCTCTGCCCC			226922
QY	1999	CysProLysGlySerPhe			2004
Db	226921	TGCCCAAGGCTCTCTTGTGATGGTTTATGATTACAATATGATATGACAACCTTGACCTT			226862
QY	2004	-----	-----	-----	2004
Db	226861	TACTCTCAGAGGAATAAAGATAGCAGCTAGAGCCAGCTTGAGCTCCGGGGCCCCACT			226802
QY	2004	-----	-----	-----	2004
Db	226801	TTCTGGAATCTATCTCCAGAGACTCAGTGTCTCACTACCTTCACTCACTCAACCATCA			226742
QY	2004	-----	-----	-----	2004
Db	226741	ACCTTAGGCTGACCAACGTCCTATATCCAGCCACACAGAGCTGACGAAATCCTGCGT			226682
QY	2004	-----	-----	-----	2004
Db	226681	GTCCTGAACAGGGCTGGGAGCTAGACTGGTGTGATTGCAATGGGGTGCCTATCTTGAGCTAA			226622
QY	2004	-----	-----	-----	2004
Db	226621	CTACTTAGTACTGAGTTACCCCACTGAAGAGAGGCCGTGCTGAGGATGTGCGTGC			226562
QY	2004	-----	-----	-----	2004
Db	226561	TGGGTACTATCCCAAGATTTCAATTAGTGTGACATGTTTCATGGGAGGAGTAGC			226502
QY	2004	-----	-----	-----	2004
Db	226501	TGAGAGTAAGGAGTAGTGTGATGACAGGCGTAGTGAATCTTGGGGCGGTGTGGCAG			226442

  

QY	2004	-----	-----	-----	2004
Db	226441	AGGTTACATAGGTCCCTTTGGCCACTGACTGATTTTGTGAGATGTTTGGCTCTGGA			226382
QY	2004	-----	-----	-----	2004
Db	226381	GCTCTGTACACTGGGCTTCTCTGTCTCTCCCTAAGCCCGCCTTCTTGGCTTCTTCTCC			226322
QY	2005	-GlyThrAlaValArgHisCysAspGlyLysArgGlyTyrLeuProProAlaLeuPheAs			2024
Db	226321	AGGACTCTGTGGCCACTGTGTATGACACAGGGGGCTGGCTCCGCCAATCTTTCAA			226262
QY	2024	nCysThrSerIleThrPheSerGlyLeuLysGlyPheAla			2037
Db	226261	CTGCACGTCACTCACTTCTCAGAGCTAAAGGCTTTGT-AAATTGACCCCTTGTCTTCT			226203
QY	2037	-----	-----	-----	2037
Db	226202	CCTCTTCTTACATACCCTGTCTTGTGAAATAGAGCTCCGGGGTTGGGATTTAGCTCAG			226143
QY	2037	-----	-----	-----	2037
Db	226142	TGTTAGAGCGCTTGCTTAGAAGCGCAAGCCCTGGGTTGGTCCAGCTCCGGGAAAAA			226083
QY	2037	-----	-----	-----	2037
Db	226082	AAAGAACAAAAAAGAAAAAATAGAGCCCGAATCCTGTTTCTCTGAGGCTCC			226023
QY	2037	-----	-----	-----	2037
Db	226022	TCTTACTGACAGAGCCACTTCTTCCCACTGATCTTGAGGCCCTTTACTGCTGAGC			225963
QY	2038	-----	-----	-----	2039
Db	225962	CCCAAGTCCCAACATCCTGTCTGACCTCTGCACTTCTGCTCTCCATCCAGGCCAGCG			225903
QY	2039	GlyeuAlaArgAsnLysSerGlyLeuAspSerGlyArgSerGlnLysLeuAlaLeuLeu			2059
Db	225902	GCTACAGGAAACCAATACAGGCTGTGACTCAGAGCGCTCCAGAGGCTGGCCGTCTCT			225843
QY	2059	ValArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrG			2079
Db	225842	GCGTAAATGCCACACAGCACACCTCTGCTACTTGGGAGTGAATGTCAAGGTGGCTTACA			225783
QY	2079	nLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnAlaArgGlyPheGlyLeuSerAlaTh			2099
Db	225782	GCTGGCACACGACTGCTGGCTCATATGAGATGCCACGGGGGCTTTGGCTGTCCGCCAC			225723
QY	2099	ArgLysPheValHisPheThrGlu			2106
Db	225722	ACAGGATGTGACTTCACTGAGGTGGGCTTGAAGTAGGCAATGGGGGCCCATGAGAG			225663
QY	2106	-----	-----	-----	2106
Db	225662	AGCTGTCTCGAGATCTTGGGGAGATCTTGAAGGAGATATCTTACTGTAGAGAG			225603
QY	2106	-----	-----	-----	2106
Db	225602	CTGGCCCAACCCCTGTACAGAGGCTCCATGATAGTCTGTAAACAGAGCCGACACAGTT			225543
QY	2106	-----	-----	-----	2106
Db	225542	GATTAAGATACCTCTCTGGGGCTGGGATTTAGCTCAGTGTGATAGGCTTAACTAG			225483
QY	2106	-----	-----	-----	2106
Db	225482	GAAAGCAAGGCCCTGGGTTGGTCCCACTCCGAAAAAAGAAAAAAGAGT			225423
QY	2106	-----	-----	-----	2106
Db	225422	ACCTCTCTTACCAAGTGTACGAATTTAGTACTTGAAGCTTGAAGGCTGACCTGAGGC			225363
QY	2106	-----	-----	-----	2106

Db 225362 AGAGCCCACTATTATACCAAGAGCTCAGTCTTGTCAGGAGCAAAAGAGCAAAAC 225303  
QY 2106 ----- 2106  
Db 225302 AGAGCCAGACATGCAGAGACATGGTGCCCTGATTAAAGTGTGAAGAAAGTTTAG 225243  
QY 2106 ----- 2106  
Db 225242 AGGCCAGTACCATGCTGGAAGTGAAGATATGATACAGAGGCTTCCAGGGAGAGACT 225183  
QY 2106 ----- 2106  
Db 225182 GTGGGAGCTGCTAAGGGAGAGACTAGTGGGCGGCTGTGAACAGGACAGGCTGAACT 225123  
QY 2107 ----- AsnLeuLe 2109  
Db 225122 TGGGATGAGGTGACCAAGACATATGCACTGTGACCATGCCATCTCCAGAACTGCT 225063  
QY 2109 uArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisTyrGluLeuIleGlnG 2129  
Db 225062 GAGGTTGGGACGGCCCTCTCGATGCGGCAATTAAGGCACTGGGAACTGATCCAGCA 225003  
QY 2129 nHnGlnGlyGlyThrAlaTyrLeuLeuGlnHisTyrGluValAlaTyrAlaSerAlaLeuAl 2149  
Db 225002 GACAGAGGTTGGCACTGCTGCTGCTGCTCCAGCACTATGAGGCTTATGCCAGTCCCTGGC 224943  
QY 2149 agLAsnMetArgHisThrTyrLeuSerProPheThrIleValThrProAsnIle----- 2167  
Db 224942 CAGAAATATCGGCATACCTACCTGAGCCCTTCACTATGTCACACCACTTGGTGA 224883  
QY 2167 ----- 2167  
Db 224882 GGTGTGCTGGGCTGGGGGCGGGGTGTGAGAGTCTGCCCATTTCAAGGAGAGCCA 224823  
QY 2168 ----- ValIleSerValAlaArgLe 2174  
Db 224822 GCCGTGCTGGTGAATGATCTTGCAACCAATTTCTTCTAGTCACTCTGTGGTGGCCT 224763  
QY 2174 uAspLysGlyAsnAspHeAlaGlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGlnG 2194  
Db 224762 GGAATTAAGGGAACTTGTCTGGAGCAAGCTGCCCGTTATGAGGCGTTAGTGGGAGCG 224703  
QY 2194 nProProAspLeuGluIuThrThrValIleLeuProGlnSerValPheArg----- 2210  
Db 224702 CCTCCAGATCTTGAAGACCAAGTCAATTTGGCAGAGTGTCTTCAAGAGTCAAGCGGG 224643  
QY 2210 ----- 2210  
Db 224642 AACGTGATGTGTAAACGCTTGGGGTGGGGACCCATGCAATGGGAGGGCCAGCTGAG 224583  
QY 2211 ----- GluThrProProValValArgProAlaGlyProGlyGlu 2224  
Db 224582 TGTAGCAGATTTACTTGACAGAGATGCTCCCATGGTGAATCTGACAGGACTGTGTAAG 224523  
QY 2224 laGlnGluProGluGlnLeuAlaArgArgGlnArgArgHisProGluLeuSerGlnGlyG 2244  
Db 224522 CCCAGGAGACTGAGAGACTGGCCCGGCGGAGGAGGCAACCACTAGTACAGGGCG 224463  
QY 2244 luAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyr 2264  
Db 224462 AGGAGTGGGCACTGTATCATCTTTTACCAACGCTGGCTGAGCTGCCCACTAG 224403  
QY 2264 sPProAspLysArgSerLeu----- 2270  
Db 224402 ACCCAGATTAAGCGTGAAGGTAAGCGGCTACGACAGGCTTAGAGGTTGGGTTAG 224343  
QY 2270 ----- 2270  
Db 224342 CAAGGAGATGTGAAGCATCTCCAGTGAATTAAGGTTGCTGTGACCGCTTGGTGACTG 224283  
QY 2271 ----- ArgValProLysArgProIleIleAsnThrProValValSer 2285

Db 224282 TGTACCTTCTGTGCATCAGAGTCCCAAGCGCCAGTCAACACACTGTGTAGCA 224223  
QY 2285 leSerValHisAspAspGluGlnLeuLeuProArgAlaLeuAspLysProValThrValG 2305  
Db 224222 TCAGTTCACATATGATAGAGAGTCTTGCCAGGGGCACTGGACAGAGCACTGACAGTGC 224163  
QY 2305 lnPheArgLeuLeuGluIuThrGlnGluAlaArgThrLysProIleCysValPheTyrAsnHis 2325  
Db 224162 AGTTCGAGCTGTGAGAGAGGAGGAGGAGGACCAAGCCCATCTGTGTCTTCCGAACCAT 224103  
QY 2325 erTlleLeu----- 2327  
Db 224102 CGATCTGTAGAGCTGCTGTACCACTTAAAGCCATGACTTGGACCAAGATGCCAG 224043  
QY 2328 ----- ValSerGly 2330  
Db 224042 CTCCTGGCTGTCACTCTTCTTCTCCGATCCCGTCTCTCTCCGAGGGTCAAGTGC 223993  
QY 2331 ThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAsnLysSerHisValSer 2350  
Db 223982 ACAGGTGCTGTGCGCCAGAGGCTGCGAGGTTGTCTTCCGTAAACGAGGACATGTACAG 223923  
QY 2351 CysGlnCysAsnHisMetThrSerPheAlaValIleAsnMetAspValSerArgArgL 2369  
Db 223922 TGCAGTGCACATATGACAAAGCTTGTGCTATGATGATGTCTCCGAGAGAGTT 223863  
QY 2369 ----- 2369  
Db 223862 GGACTCCAGGGGTAGCCCAAGAGCAGCGGTGGGACCAAGTCAAGGGGCTCAAGCCCT 223803  
QY 2370 ----- AsnGlyGluIleLeuProLeuLysThr 2378  
Db 223802 GCTCTTTTGAACCCCACTCACTCACTCCAGAAATGTGAGAAATTTGGCACTGAAGACC 223743  
QY 2379 leuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeu 2398  
Db 223742 CTGAGTATGTGACCTTGAGACTCACTAGTGTGCTGATATATCACTTCTCTTCTCC 223683  
QY 2399 ThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAla 2418  
Db 223682 ACCCTCTTTCGAGCCCTTGCTGCCAAGCAGATGAGATCCAGCGAACCTTCACTGCTGC 223623  
QY 2419 leuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuPro----- 2436  
Db 223622 CTGGGCTGGCCCACTGCTCTTCTCTGGGAGTCAACAGGCTGACCT-CCCTGTAAG 223564  
QY 2436 ----- 2436  
Db 223563 ATACTTCTCTCCAGAGACTTCCCAACTTCAAGCCCCACCAAGGCTCATAGTCTCCC 223504  
QY 2436 ----- 2436  
Db 223503 TGTATGACCTTGTGCATCCCTGCTCCACAGCATCATGATGAAGGCCCATTCAG 223444  
QY 2436 ----- 2436  
Db 223443 TACTGCCCAAGTTTGCCCTTCTCTCAACACAGAGCTTCCCACTCCCTGGGGCA 223384  
QY 2437 ----- PheAlaCysThr 2440  
Db 223383 CTCCTTGATTAACGTCTGTGCTCTGTATCATGTCTCGGCGCTCAAGTTCTTGACCA 223324  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTyrAlaLeuGln 2460  
Db 223323 GTCAATGGCACTGTCTGCACTTCTGTATCACTTTCACCTTCTCTGGGCTGTGCTGAG 223264  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
Db 223263 GCTTACACCTGTACCGGGCGTCAAGAGGTGCAGAGCTAAATGCCAGCCCATGTGCT 223204  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr----- 2494  
Db 223203 TTCTACTATACGTGTGGGCTGGGCGCTTCTGCTTTCATCAACAGTATCTGTCTCTTTC 223144

QY	2494	-----	2494
Db	223143	CTGACCTAGAGCTCCCTGTGTGTAAGGCTGGGGTCACTTCATGATGATACCCCAAA	223084A
QY	2494	-----	2494
Db	223083	CCATGAATTCCTGAGAAATCCCTTCCCTTACTTGAGACAGCTGATTAAGTGTCAAG	223024A
QY	2494	-----	2494
Db	223023	GGCTAAGCTTCTCTTGAGAGGGAGTGAGACAGCGTCTCTGTGATGCTCCCACT	222964A
QY	2494	-----	2494
Db	222963	GGCTCTGTGAACCTTGATGATCATATCTTGCTCAACCTGTAGCACCTGTCTTGCTG	222904A
QY	2495	-----	2495
Db	222903	TCGTCTGTGCTCAGAGTCTGGGTGTGGGCTTGATCCGAAAGCTATGAAACCTTGAC	222844A
QY	2510	PheCvTTPLeuSerLLeTyrAspThrLeuLLeTyrSerPheAlaGlyProValAlaPhe	2529
Db	222843	TTCTGTGTGCTCTCCACTTACGATACGTCATCTGAGATTTCCTGGAACAAGGGCTTT	222784A
QY	2530	AlaValSer-----	2532
Db	222783	GCTGTTTCGTTAAGTGTCTTGAAAGCAATTTAGGGTTAGCAGATGCGAGGGTCTTACAG	222724A
QY	2532	-----	2532
Db	222723	TTCCCTTTAAGGAAGAGTGAAGCTCTGGGCCCCCTGGGCAAGAAACTGTCTTG	222664A
QY	2533	-----	2542
Db	222663	CTCTGTGCTCTCATGTGCCCACTGCAGATGATGATGTCTTCTCTGATATCCCTGTGGCC	222604A
QY	2543	ArgAlaSerCvAlaAlaGlnArgGlnGlyPheGlnLysLysGlyProVal	2559
Db	222603	CGAGCTCTGTGTGTGCCAAGTCAGAGGCTTCGAAAGAAAGGCCCTGTGTGATGACTG	222544A
QY	2559	-----	2559
Db	222543	GCAAGTGTGGGCAAGGGCGGGTGTGTATAGAGAGCATATGGCGGGTGTGTAT	222484A
QY	2560	-----	2565
Db	222483	AGAGAGCATCATGGCGGAATCATATATGCCCTGACCTCCAGCTCAGGCTGTGGCTCTTC	222424A
QY	2565	PheAlaValLeuLeuLeuLeuSerAlaThrTTPLeuAlaLeuLeuSerValIAsnSe	2585
Db	222423	CTTCACTGTCTCTCTGTGTGTAGTGTCCACATGGCTGTGGCACTGTCTGTCAACAG	222364A
QY	2585	TAspThrLeuLeuPheHisTyrLeuPheAlaThrCvAsnCvAla-----	2600
Db	222363	CGAACACTCTCTTCCACTACCTTTGTGGGGCTGCAATTGTGT-CAAGTACTTAGCC	222305A
QY	2600	-----	2600
Db	222304	CCGCCCTACTGATGAGAGAGCATATCTGGGGGTGGGTGTGTCGTAACCATAGGCTGTG	222245A
QY	2601	-----	2618
Db	222244	TGCGTCCAGGGGCCCTTTCATCTTCTCTCTTCATACGTGTGTCTCAAGCAAGAGGTCCGGAA	222185A
QY	2618	AlaLeuLysLeuAlaCvSerArgLysProSerProAspProAlaLeuThrThrLysSe	2638
Db	222184	AGCACTCAAGTTTCGCTCTGAGCGGAAGCCAGCCCTGACCTTGACCTTGAACCTAATGTC	222125A
QY	2638	ThrLeuThrSer-----	2642
Db	222124	TACCTTGAACCTTC-GGTGAAGGGGCCGGGGGTCTGAGCTCAAGAGCGCTCTTGTGTAGAAGG	222066A

[illegible]

```

Db      220985 GCTTCAACCACTCCTACAGACAC-----GAGGAGAGAGAGAGAGGCGCC 220938
Qy      2755 PheProGluGluGluGluGluTPAPSerLeuLeuGlyProGluValaGluLeuProLeu 2774
Db      220937 TTCCTGGGAGAGAGAGGCTGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 220878
Qy      2775 HisSerThrProlys----- 2779
Db      220877 CACAGTACCCCA--GGGTGGGCGCAGACACAGGCTGACGCTTGGGGGGGCGACAGGAA 220819
Qy      2779 ----- 2779
Db      220818 GCCTAAGAGGCTAGGCTTCTTGGCTATCTCTGCTCTACCTCTCTGGAAGCTCATAGGCTTC 220759
Qy      2780 -----AspGluGlyProGlyProGlyValaAlaProThrProGlyAspPheGly 2795
Db      220758 CTTTCTTTACAGATGGGGCTCCAGGGCTGGAGAGGCTCTTGGCGCAGAGACTTTGGG 220699
Qy      2796 ThrThrAlaGluSerSerGlyAaenglyAlaProGluGluValaGluGluAengly 2815
Db      220698 ACCACAACAAGAGAGATAGTGTAGTGGGCCCCCTTGAAGAGGCGCGACGGAGATGGA 220639
Qy      2816 AspAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySerSerAlaGlnProHis 2835
Db      220638 GAGGCTCTAAGCTGGAGAGGCTCTGAGGACCCCTTCCGGGCTCTTCTACCCAACTCAG 220579
Qy      2836 LysGlyLe----- 2838
Db      220578 AAGGTTA-GTGAAGAGTCTCTCTCATCTGCTTCTGCTCTGCTCTTCTCTGGAATTT 220550
Qy      2839 -----LeuValysLysCys 2843
Db      220519 TAGAATGCTTCATCTCATGTGATGCCACACCTCTTACCCAGGACATCTCTCAAGAGAGTGT 220460
Qy      2844 LeuProThrLeuSerGluLysSerSerLeuLeuArgLeuProLeuGluGluGlyThrGly 2863
Db      220459 CTGCCACCACTACGCAAAAGAGTACCTCTTAAGTGTGCTTGAAGAGCGGCGACAGGG 220400
Qy      2864 SerSerArgGlySerSerAlaSerGluGlySerArgGlyValProProProArgProPro 2883
Db      220399 TCTTCCCGGGGCTCCACCGCATGAGGCGACAGACGACCTCTCTCCACGCCACCA 220340
Qy      2884 ProArgGlnSerLeuGluGluGluGluValaMetProIleAlaMetSerIleLys 2903
Db      220339 CCGCGCCAGTCTCTCCAGAGACGTGAATGGGTATGGCCATCGCATGACATCAAG 220280
Qy      2904 AlaGlyThrValaAspGluAspSerSerGlySerGlu 2915
Db      220279 GCAGGCACTGTGATGAGAGACTCTTGTGCTCCAG 220244

RESULT 15
AC113756/c 248059 bp DNA linear HTG 15-NOV-2002
LOCUS Rattus norvegicus clone CH230-195D20, WORKING DRAFT SEQUENCE, 7
DEFINITION unoriented pieces.
AC113756
AC113756 GI:25006575
HTG: HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 248059)
Nuzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrechts, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, K., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, R., Blyth, P., Brown, M.,
Byrant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacho, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

```

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De And, C., Dederich, D., Delgado, O., Denson, S., Dexam, C., Ding, Y., Dint, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladik, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louisedge, H., Lorado, R.T., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mamoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mawhinney, S., McLeod, M.P., McNeill, T. Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelamen, O., Okwomu, G., Olampusaogon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, R., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S., Sander, M., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisom, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemari, C., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wiczysk, R., Woodem, H., Wotley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 248059)  
Worley, K.C.  
Direct Submission  
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 248059)  
Rat Genome Sequencing Consortium.  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23269994.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

```

----- Project Information
Center project name: GSES
Center clone name: CH230-195D20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226746 bases at least Q40
Consensus quality: 228921 bases at least Q30
Consensus quality: 230687 bases at least Q20
Estimated insert size: 231376; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7257: contig of 7257 bp in length
* 7258 7357: gap of unknown length
* 7358 54651: contig of 47294 bp in length
* 54652 54751: gap of unknown length
* 54752 58145: contig of 3394 bp in length
* 58146 58245: gap of unknown length
* 58246 75921: contig of 17676 bp in length
* 75922 76021: gap of unknown length
* 76022 24564: contig of 16943 bp in length
* 24565 24566: gap of unknown length
* 24566 246797: contig of 1233 bp in length
* 246798 246897: gap of unknown length
* 246898 248059: contig of 1162 bp in length.
*
* Location/Qualifiers
*     1..248059
*     /organism="Rattus norvegicus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10116"
*     /clone="CH230-195D20"
*     54752..55991
*     /note="wgs contig"
*     58246..59729
*     /note="wgs contig"
*     74011..75921
*     /note="wgs contig"
*
BASE COUNT  61509 a 54522 c 53670 g 62050 t 16308 others
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 248059
Score: 11892.00 Matches: 2783
Percent Similarity: 38.18% Conservative: 47
Best Local Similarity: 37.55% Mismatches: 77
Query Match: 76.50% Indels: 4515
DB: 2 Gaps: 33
US-09-916-849a-3 (1-2923) x AC113756 (1-248059)
QY 5 A1aThnglyValProleuProthProProProleuLeuLeuLeuLeuLeu 24
DB 40197 GTGGCCAGGCGCCCTCCCAACGCGCACTTGATGTTGCTACTGCTGCTG 40138
QY 25 LeuProProProleuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 44
DB 40137 CCGCGGTGGCGCACTGGAATCAAGTGGGCGCTGCTGCTGCTGCTGCTG 40078
QY 45 ArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCysProSerSerAlaSer 64
DB 40077 CGCGGCTCTCTGCGGCGCTGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTG 40018
QY 65 LeuTyrLeuTyrThrSerArgCysArgAspAlaGlyTyrThrGlyLeuThrGly 84

```

```

DB 40017 CTGCGCTCTTACACGAGCGCTGACAGATTCGGCGATTGATGACCGGCACTG 39958
QY 85 ProHisAspGlyLeuArgValTyrCysProGluSerGluAlaHisIleProleuPro 104
DB 39957 CCCCATCAAGATGGCTTGAAGGTTTGTGTCTCCAGATACAGGGGCTCATATCCCTCTTCCG 39898
QY 105 ProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGlyValHisLeuSer 124
DB 39897 CCATCCTCTGAAGGCTGCGCCCTGAGAGCTGTGCACTCTGGGTATCGAGAGACCTTTCT 39838
QY 125 ProGluGlyValLeuThrLeuProGluGluHisProCysLeuValAlaProArgLeuArg 144
DB 39837 CCACAAAGGACGCTGACCTGCTCCACAAACACCTTCTTAAAGGCCCAAGGCTCGA 39778
QY 145 CysGluSerCysLeuLeuAlaGluAlaProGlyLeuAlaGlyValGluArgSerProGlu 164
DB 39777 TCCAGTCTTTCAGCTGCGACAGGCGCCAGGGCTCAGGGCTGGGAAAGATCAGAGG 39718
QY 165 GluSerLeuGlyValArgArgCysArgAspValLeuThrAlaProGluPheGluProPro 184
DB 39717 GAATCTATGGGTGGCGCAGGAAAGATGTAATACAGCTCCCAAGTCCAGCTCC 39658
QY 185 SerTyrGlnAlaThrValProGluAsnGluProAlaGlyTyrProValAlaSerLeuArg 204
DB 39657 AGCTACCAAGCCACAGTCCCGAGAACAGCTGCTGTACTCTGTGCTCTCTTAAGA 39598
QY 205 AlaIleAspProAspGluGlyValAlaGlyValArgLeuGlyTyrThrMetAspAlaLeuPhe 224
DB 39597 GCCATATATCCGATGAGGAGGAGGCGGTGCACTTGATGATACCATATGATGCTCTTT 39538
QY 225 AspSerArgSerArgGluPhePheSerLeuAspProValThrGlyAlaValThrAla 244
DB 39537 GATAGCCCTCCATCATTTCTTCTCTTGACCAATCATCTGGGTGCTCACCACAGCT 39478
QY 245 GluGluLeuAspArgGlyThrThrSerThrHisValPheArgValThrAlaGluAspHis 264
DB 39477 GAGGAGCTGAGTCCGAGAACCAAGAGCAACCATATCTTCAGGGTCAACGACAGATCAT 39418
QY 265 GlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAsp 284
DB 39417 GGTATGCCCCGACGAGTCCCTTGGCCACATCTATCTTATGATCAGACCAATATAT 39358
QY 285 HisAspProValPheGluGluGluGlyTyrIleGluSerLeuArgGluAsnLeuVal 304
DB 39357 CACACCTCTGTTTGGAGCAGCAAGATATACAGAGAGCTCAGGAGAACTGGAGGTT 39298
QY 305 GlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIle 324
DB 39297 GGTATGAGGTGCTTACAGTCAAGGCGCACCGATGCGATGCCCTCTTAATGCCAACATT 39238
QY 325 LeuTyrArgLeuLeuGluGlySerGlyGlySerProSerGluValPheGluIleAspPro 344
DB 39237 CTGATCCCGCTGCTGAGAGGGCTCGAGAGGAGCCCTCAGAAAGCTTTGAGATGATCTT 39178
QY 345 ArgSerGlyValIleArgThrArgGlyProValAspArgGluGluGluSerTyrGln 364
DB 39177 CGCTCTGGGTGATCCAAACCGCGCGCCCTGATGCGGAAAGATGATCTTACAA 39118
QY 365 LeuThrValGluAlaSerAspGluGlyArgAspProGlyProArgSerThrThrAlaAla 384
DB 39117 TTGATGAGGAGGAGATGACAGGATCGAGGCGCACCGAGCCACAGAGTCCACAGCTGTC 39058
QY 385 ValPheLeuSerValGluAspAspAspAspAspAlaProGluPheSerGluValArgTyr 404
DB 39057 GTCTTCTGCTGTGTGGAAGATGATTAAGACACCCCTCAGATTATGAGAGACGCTAC 38998
QY 405 ValValGluValArgGluAspValThrProGlyAlaProValLeuValThrAlaSer 424
DB 38997 GTGATTCAGGTGCGGAGAGTGAACCCAGAGAGCCCGCTTCTCCAGATCAGCGGCTCG 38938
QY 425 AspArgAspArgGlySerArgAlaValAlaValHisTyrSerIleMetSerGlyAsnAlaArg 444
DB 38937 GACAGAGCAAGGCGACCAACGCTGTGATCATATGATCATAGCGGCAAGCTCGG 38878

```

QY	445	GLYGLNpheThyLeuAspAlaGlnThrGlyAlaLeuAspValValSerProLeuAspTyr	464	QY	805	ValAsnAspValAlaAspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerVal	824
Db	38877	GGAACATTCCTATTCGATGCTCAGACTGAGACCTCGATGTGGTAGGCCACTCGACTAT	38818	Db	37797	GTAATAGACGTGAATGACATGCCCCCAAGTTCTCAAGATTCCTTACAGGGAGTGT	37738
QY	465	GIuThrThyLeuGluTyrThrLeuArgValArgAlaGlnAspGlyValArgProLeu	484	QY	825	TyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSer	844
Db	38817	GAGACAAACCAAGATATACACTCGCGATCCGGGCCAAGATGTGTGGCCGCTCCACTT	38758	Db	37727	TATGAGATGTGCCACCTTCACCAAGCGTCTCGAGATCTCAGCAGCCGATGCCACTCC	37678
QY	485	SerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIle	504	QY	845	GlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGlyAspGlyAspPhe	864
Db	38757	TCCAAAGTCTCCGGTCTGATACCGGTGACGTCTGACATCATATCAACACGCCCCATC	38698	Db	37677	GGTCTGAACGGGAGGTTTCTTACCTTCACAGAGGGGAGCATGAGATGTGTACTTT	37618
QY	505	PheValSerThrProPheGlnAlaThrValLeuGluSerValProLeuGlyTyrLeuVal	524	QY	865	IleValGluSerThrSerGlyIleValArgThrLeuArgGlyLeuAspArgGluAsnVal	884
Db	38697	TTTGTACGACACCCCTTTCAGGCCACGGTCTGAGAGTGTACTTTAGGCTTACCTGGTT	38638	Db	37617	ATTGAGAGTCACATCAGGCATGTGGTGCACATCGCGAAGCTGTGATGTGAGATGTG	37558
QY	525	LeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyrArgLeu	544	QY	885	AlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProProAlaArgThrPro	904
Db	38637	CTGCATGTCCAAAGCGATTGATGTGATGCGGAGAGAACGCCGCTCGAATACAGCTTG	38578	Db	37557	GCCCAGTACATCTTGAGAGCGTACCGGGTGACAAGGGGATGCGCGACAGCCACACCC	37498
QY	545	AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyTyrGlyTyrIleSerVal	564	QY	905	MetGluValThrValThrValLeuAspValAsnAspAsnProProValPheGluGlnAsp	924
Db	38577	GCTGGGGTTGGGACGACTTCCCTTCACATTAAACAGCGCACAGGCTGATCTCGGTG	38518	Db	37497	ATGGAAGTGAAGTACTGTACTGTGATGTGAATGACAAATCCCCCTGTTTGAACAGGAT	37438
QY	565	AlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAsp	584	QY	925	GluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaValAlaArgValThr	944
Db	38517	GGCGCGGATTTGGACCGGAGAAAGGTTGATTTTCAACGCTTCGAGATGAAAGCCAGGAC	38458	Db	37437	GAGTTTGAATGATTTTGTGGAAGAGAACGCCCATCGGACTGTGTGTGGCCGGGTCCA	37378
QY	585	HisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAsn	604	QY	945	AlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIleValGluGlyAsn	964
Db	38457	CACGGACCCCGACACTCACCGCTCAGCGATGTGATGATTAACATCTCGATGTCAAT	38398	Db	37377	GCCACCGACCCAGAGGAGCAACATGACAGATCATGTACAGATGTGAGGGCAAT	37318
QY	605	AspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaIle	624	QY	965	IleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAlaLeuValAspLeu	984
Db	38397	GACACACACCAACCTTTACGACGCGAGATACAGGTGGCTCAACAGAGACGGGGCT	38338	Db	37317	ATCCCTGAGTCTTTCAGCTGACATCTTCTCGGGCGAGCTAACGGCCCTGTGATATTG	37258
QY	625	ValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleThr	644	QY	985	AspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThrSerAlaProLeuVal	1004
Db	38337	GTGGGACCAAGTGTGTGACGGTGTGACGGCTGGAGTCAAGATGCTCACGTGCTCAATAC	38278	Db	37257	GACTACAGAGACCGGACCGATATATCTCGTCAATCCAGGGCTAGCTGCTCTTGGTG	37198
QY	645	TyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGly	664	QY	1005	SerArgAlaThrValHisValArgLeuLeuAspArgAsnAsnProProValIleuGly	1024
Db	38277	TACCAAGATCACAGCGGCAACCCCAACCGCTCTCATACCCGCCAAGCGGGGCT	38218	Db	37197	AGCAGGGCTACGTCCAGTCCGGCTCTCGACCGCATGTATTAACCCCGCAGTGTGGGC	37138
QY	665	GlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGlnTyrValLeuAla	684	QY	1025	AsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSerSerPheProGlyGly	1044
Db	38217	GGCTGTGGTCTCCCTGACCTTACCGCTGACATCAACCTGAGGGGACGATGTGCTGGCC	38158	Db	37137	AACTTTGAATTCCTTTTCAACACTATGTCAACAAACCGCTCGAGCGCTTCCCTGGGGGT	37078
QY	685	ValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAsp	704	QY	1045	AlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeuThrTyrSerPhe	1064
Db	38157	GTGACTGTCTCGGATGGGACAGGAGGACGACACGCTCAATGTGTGTGATGTCACTGAT	38098	Db	37077	GCTATAGGCGCGGTCGCTGCCATGACCCCGATATCTCAGACAGCGTGAAGCTTTT	37018
QY	705	AlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsnValLeuGluAsp	724	QY	1065	GluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGlyGluLeuLeuLeu	1084
Db	38097	GCCAAACACCATGTCTCCGTCTCCAGACTCCCATATACGCTGATATCTTATATGAAGC	38038	Db	37017	GAGCGAGAAATGAATCAAGCTGTGCTCTCATGTGCTTCCACCGGAGACTGAGACTG	36958
QY	725	ArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsn	744	QY	1085	SerArgAlaLeuAspAsnAsnArgProLeuGluIleAlaIleMetSerValIleVal-----	1102
Db	38037	CGGCCAGCGGACCACTGTGTGTGATCAGTGCTTACAGAGAGGACACAGGGAGGAAT	37978	Db	36957	AGCGGGCACTGACACAAACCGGCTCGGAACCATCATGATGTGTCTGT-CTCAGG	36899
QY	745	AlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspThr	764	QY	1102	-----	1102
Db	37977	GCCCGAATCACTTATATGAGGACACATCCCTCACTTCGCGATCATGACAGCACT	37918	Db	36898	TAAAGAAATGCCAGTGAATGTGGGTGGGAATATGCTTTAGGAGAGTCTTGGAAAGCC	36839
QY	765	GlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAla	784	QY	1102	-----	1102
Db	37917	GGGGCGGTACCAACCGAGGCTGAGCTGACTAGAAAGATCAAGTGTCTTACACCCCTGGCC	37858	Db	36838	ACTGAGGTGTGGGCTTCCAAAGAGGGGCTGTGAACCAAGCTCTCAACCTGTGATAG	36779
QY	785	IleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyrLeuGluIleLeu	804	QY	1102	-----	1102
Db	37857	ATCACCGCTCGGGACAAATGGCATCCCCCAAGTGGGACAACTATCTGAGATCTCTG	37798	Db	36778	CTCAAGATTGTGATCAAGGAACCTGGCTGAGAAAGAAACCCAGGAATCTTGGCAGCT	36719
QY				QY	1102	-----	1102

Db	36718	GGCGCAGGCCCCCGCCTGCTGCCCCAGGCTGAGCTGAGCTCTTGGTTGGCTTGCCTCCCTG	3655
OY	1102	-----	1102
Db	36658	CCTACACAGACTGTGACGTGCTGGGGGAAGCCTTGTGAGGCTGCCCTTGGCTTGGCTTGC	36599
OY	1102	-----	1102
Db	36598	AGAAATTGTGAAAAAAGGGGCGCTGAGAACCTTGAAAGGCCAGAGCTGGAGAG	36539
OY	1102	-----	1102
Db	36538	GAGGAGAGGGGAGCTTGGGAGAGAGAAAAAGGGAGGGGGAGGGGAGTGGAAAGACTTGGCT	36479
OY	1102	-----	1102
Db	36478	GGCAGGGGATGGGGTGGAGGGGAGGCAAAATCGGTTCTCAGAACCTGTCACTCCCTCATG	36419
OY	1102	-----	1102
Db	36418	GCACCTGTGTATCTCAACAGCCTTTGGCAGGTCCAGGGCTACAGAGACACAGAGATTTC	36359
OY	1102	-----	1102
Db	36358	CACCAAGGGGTCAGAGACCTGCGGCTCTCTCTCCCACTCGGTTCTCAGTAGAGCT	36299
OY	1102	-----	1102
Db	36298	GCAGAAACTGACTTGGACTTGTCTAGAGCATTCCTTTCCCTCAGGCGAAAGAGCT	36239
OY	1102	-----	1102
Db	36238	GCTTGGCTCCCTTCAGAGCTTCAGGCAAGAAATGAGTCAAGCTCATTTGTCTTGT	36179
OY	1102	-----	1102
Db	36178	CCAGACAGACTCACTGCCCTGAGTTCTTGGTTGGGCTTGGCTTGGGGAAAGGCTGCA	36119
OY	1102	-----	1102
Db	36118	ACAAAGCCAGGACAGATGACTTTTGGGCTCCCCACCCCCGGTGGCTTGTCTTGGGG	36059
OY	1102	-----	1102
Db	36058	TGCTAGGTTCTTCTGCTGACCATGATGCGATGAGCTTCAGAGAACAGACAGGGGTTCA	35999
OY	1102	-----	1102
Db	35998	GAGTGCACTAGGACTACTGCTCTTTGAGGGCAAGACTCCAGGGCCCTGACTTGGTGGAG	35939
OY	1102	-----	1102
Db	35938	AGAGACAGATTTCAGGGGCGTGGAGCTGGCCAGAAACCGACTGGGCATTCCTGCA	35879
OY	1102	-----	1102
Db	35878	GCTTGGTCAGCGCTCAGCCCCCAGACAGACAGACTGGGAGGGAGACAGAGACTGGGT	35819
OY	1102	-----	1102
Db	35818	GATAGTATTCTATCCCTTCTGTCACTCTGGAAATGAAAGCAAGCGGGAAGGGGACAG	35759
OY	1102	-----	1102
Db	35758	CTCAGTAGCAAGACCTTTCCTGAGTCTGGAGTTGGGGCTGGAGAGAGACTGTAAAT	35699
OY	1102	-----	1102
Db	35698	TTTTCCAGGGGTGAGACTATGTGGCTTACTCAGGCAAGTCTTACATCATGAGGAGGGGAA	35639
OY	1102	-----	1102

Db	35638	GATGGCCCTGGCCCCCTTTCCTGGCTCACTTCTTACTGACAGACACTTCCTGATCCAC	35579
QY	1102	-----	1102
Db	35578	CCCTCAGGTGCTGCAAAGACAGAGAGAGAGCTCTCCAGGCTATGCTTGTCAAGC	35519
QY	1102	-----	1102
Db	35518	TGTAAACCACTGAAGCTTAGTTCCTTCAGAGAAAGAAAGCTAGAGACCTCACC	35455
QY	1102	-----	1102
Db	35458	TGAGCCTGGACAGAGAGAGAAAGAGGCTCTCCCTCCGCTCTCCATCCAGAGCTT	35399
QY	1102	-----	1102
Db	35398	CAGTGTGAATCCGGCTAATTCCTTAGGCCCCGTGTGAGAGACAGCTCTGCTTACCTTT	35339
QY	1102	-----	1102
Db	35338	GCTTCTCACCAGGGCACAGGCCCACTACAAACCTTAGAGATAGAGGAAAGGCTTG	35279
QY	1102	-----	1102
Db	35278	GCCGACGTAGGGCACTGTGGCTGCTTCTGACCCTTGTCCCTCCCTCCATCAGCCCT	35219
QY	1102	-----	1102
Db	35218	GACCTTAGAGTCACTACCAAGGCAACCCAAATTCAAGAGGCTGGGTGAGAGACT	35159
QY	1102	-----	1102
Db	35158	GAGCCCACTGCTCTTGTCTTCTCTCTGTGTTCTTGTCTCCTTCCCTCCACACC	35099
QY	1102	-----	1102
Db	35098	CTGCCAATGGGAAACCTCCTGCACCTGCTGTGTCACTTCTCCAGCTTCTGTGCG	35039
QY	1102	-----	1102
Db	35038	CTTTCCTTCAAGCTTCTCCTCACTGCTGGCTCTTGTCTTTCCTGTTCTCTATGTTT	34979
QY	1102	-----	1102
Db	34978	AGATCCGTCTTCTCTCTCACTCTGCTGAGCTGCCAGCCTGGGCTCCATGGCTC	34919
QY	1102	-----	1102
Db	34918	CTGAGCCCAATGTTCGGGCTGCTATCTCTTACAGACACAGAGAACTGCCAAGG	34859
QY	1102	-----	1102
Db	34858	CTGTCTACCTCCAGACAGAGAGTCTTATCAGAAAGACAAAGTTGGAGGCTGACT	34799
QY	1102	-----	1102
Db	34798	CTGGCTTCTCCTGTGAGCTGGCTGAGGCTTAGACCCAGGTGATATTTGCTGGGTCA	34739
QY	1102	-----	1102
Db	34738	GAGCTTACCCAGCTCTCTTCTTAAACAAAGTCAGGCCACTCTATTTCCAACTGCTT	34679
QY	1102	-----	1102
Db	34678	TCGGAAAGAGACATATGAGAAAGAGAGACAAATGAGTTGTGTGAGATGGGACA	34619
QY	1102	-----	1102
Db	34618	GTAGCAGAGGGCAGGTGGCCACAGATGTGGGCCAGATGGGAGACAAAGACATGCC	34559
QY	1102	-----	1102
Db	34558	GAGCAAGGGCACAGTCAAGGCTTGGCTGTGAACCAATGCGGAGATGTGGCTTCCAG	34499



QY	1102	-----	1102	-----	1102
Db	34498	GACCAGAGATGGCAAAAGATTCTCTTGCTTGCTGTTCAAGAGCACTGTTGTAGATCC	34439		
QY	1102	-----	1102	-----	1102
Db	34438	CTGTGTCTTTCTGTCAAGAGGGAACATCCATCCAGAAAAACATGGGCTGTAACTTGAG	34379		
QY	1102	-----	1102	-----	1102
Db	34378	CTCCTGAGCCCAAGTGCAGGCTTCTGACAGAGAGCAAAAGAGTCTGATCACATGTTGAC	34319		
QY	1102	-----	1102	-----	1102
Db	34318	CCATGGCAGAGCTGTGAAGGGCTTAGAGGGCTGTGACTTGTCTGTCTTGTAACCTGGGCAAA	34259		
QY	1102	-----	1102	-----	1102
Db	34258	GCAGAGAGACAGAGGTGTAGACAGCACTCATGCGACAGGCTTGTAGAAACAAGACATCA	34199		
QY	1102	-----	1102	-----	1102
Db	34198	CAACAGAGAGTTGAAGGGCAGGTAGGCGGCAAGCCAGGGCAACACTGAGATGGGTACC	34139		
QY	1102	-----	1102	-----	1102
Db	34138	TGCCTTCAGTATGCCCAAGCCACTTGAAATGTGACATATACAGAAATAGGGTCCATATCG	34079		
QY	1102	-----	1102	-----	1102
Db	34078	CTTTTCTACTTCTGCTTCAGACAGCACTGCTCCAGTCCATCCACCTAGATGTATATT	34019		
QY	1102	-----	1102	-----	1102
Db	34018	TTTGCTTCATGTGAGAGTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	33959		
QY	1102	-----	1102	-----	1102
Db	33958	CTCTCTCTCTCTGATGT	33899		
QY	1102	-----	1102	-----	1102
Db	33898	TTTCTGTGTACTCTGTCTGTCTCTTAAGATTGTCTGTGTGTCAAGGCTGCCCTGAATCC	33839		
QY	1102	-----	1102	-----	1102
Db	33838	AGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	33779		
QY	1102	-----	1102	-----	1102
Db	33778	AGTTGCTGCCACGTTTAAAGATGCTTTATGTAAAGGATTTCTATCATGTGTGTAGT	33719		
QY	1102	-----	1102	-----	1102
Db	33718	TTTTAACTAGAAATACCCCTTATCACCCCTTCTAAGCGTTTCCCTTCTTTGCTCCAGT	33659		
QY	1102	-----	1102	-----	1102
Db	33658	TTTCAATTTTAACTTAGTCAATTTGCCCTGTGTATGAACAATTCAGGTGTGACTT	33599		
QY	1102	-----	1102	-----	1102
Db	33598	CTGAGCTTAGGTGCGTGGGTGTGAGTTGTGGGCAACCGTGAATGTGTAGCTTTGTGAGTG	33539		
QY	1102	-----	1102	-----	1102
Db	33538	AGTGCCTGCTAGGGGTATGTGACAGGACACTTTGGGTACAACTTGTGAGCAAGTGA	33479		
QY	1102	-----	1102	-----	1102
Db	33478	ATGATGAATTCCTTAAGAGGCCCTTAGTGTCCAGAGAGAGAGACGCCCAATGGGGCTCAGGG	33419		

QY	1102	-----	1102	-----	1102
Dp	33418	TGAAAGCTGCCGCTCCTTTCTATTTTCTACAGATCCACCCCCCAACCCCCACACCT	33355	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33358	TTCTCAGCTTGGGGACATGGGCGGTGTGACACTGGCCGAGGTGGCTCAGCTGATTTTC	33299	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33298	TTGGGGTGGGGAGCCAGGCTGCCAGAGTTGGATGGAGAGAGGCTCGGTAGAGGTG	33233	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33238	TTTGCAATGGGCTTAGCAGGTTTGCACGAAGCAGTTATGGAGCTGATGCAAGGGAAAC	33177	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33178	AGAGTCAGTGAAGACATGTTTTCAGAACGCAATTCTGTCCACACCAACCCGACAGCA	33111	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33118	CGGTGAAGGCGGGGTTATCGTGGGGTTAGGTGTACGGGGCAGAGTCTCTCTGTCC	33055	-----	1102
QY	1102	-----	1102	-----	1102
Dp	33058	TCGTGGCCCTCTAGAAAAGAAAGAACAAAGACCGTTTTGGAGCAGGGAAATCCAGG	32999	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32998	CAGTCGAGGGGACAGGTATTACTGATTAAGGTGAAAAGCAGAAACCTGCAGGCAGTTA	32933	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32938	GTTTCAGGCTTCATGCCAAGAGGTGGGGAAAACAAAGAAACGGTTTCGCAATGCTTC	32877	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32878	TTTTCTTTAAGCATAATGACGTAGCCCTCAAGAGAGACTTGTCCACAGTGACACA	32811	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32818	GCTCATTTGGCCACCTTTATTCGCAATAGAAAAGTTTCTTGGGTACTTGCAGATAT	32755	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32758	GAGGTAGATTTTAGCCTGCTGACCTTCTATTCGAGATGCTCCTTAATGATGCTCC	32699	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32698	CTCTTGATCTCTCCTCCTGACCCCTCCTCTTCTTAGTACTCTTCTTCTGTGTCT	32633	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32638	CTCCTCCCGGCGCTCCTCTTGTACTCGGTTTTCTAGTGTCTTACTCTGATGCTC	32577	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32578	TCCTCTCTGTGTCTCCTCTTCTCAATACTTCTCCTGTGTGCTCCTCCTGCAGAC	32511	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32518	ACTCCCTCCTCTTGTCTCGGTTTTCTTGACTCCTCCTGACAGAGCTACTCCCTCTTC	32455	-----	1102
QY	1102	-----	1102	-----	1102
Dp	32458	GCTCTAAGCAGGAGCTTGGCGCATCTGCTTGCCTAACCCCAAGGCTAGTTGATTGG	32399	-----	1102
QY	1103	-----	1118	-----	1118
Dp	32398	TGTTGCTTCTCTTCAAGATGCTGTCCACAGGTGACAGCCAGTGTCTCACTCCGTGTACC	32333	-----	1118
QY	1119	11e1e1eThzApGluNectLeuThzH1eSer11eThzLeuArgLeuGluApMetSerPro	1138	-----	1138





```

Db 27938 GTCCTAGTGGCAGGGGTATGGGTAAAGAAACAAGACTTAATTCCTGTCCGTGTC 27899
QY 1516 -----SerLeuAsePLeuThrGlyProLeuLeuGlyGly 1527
Db 27898 TTCTTCATACCTTGTGTGTCCTCCAGAGCTCTGTGACCTGACAGAGGCCCTTGTCTGGGTGG 27839
QY 1527 yValProAsePLeuProGlySerPheProValArgMetArgGlnPheValGlyCysMetArg 1547
Db 27838 GGTGTCAGATCTGCTCCGAGAGTTCTCTGTCCGAATGCGGCACCTTGTGTGGGCTGCATAGA 27779
QY 1547 GAsnLeuGlnValAsePLeuArgHisAlaSerMetAlaSerPheLeuAlaAsnArgGlyThr 1567
Db 27778 GAACCTCCAGGTGAGTAAGCCGGACGTCGACATGCGCCACTTCATGCGCAACAATGGCAC 27719
QY 1567 rValProGly----- 1570
Db 27718 TGTGCTGCG-TATGAGAGACTGGGGGTACGGGCAAGGCTGGAGCCAAATGCGCATGAGGA 27660
QY 1570 ----- 1570
Db 27659 TCGGTTGTGATGACACAGGGAATGCCAGGCTGCGGTGTTCTTCCGTCACGTGCTGTGGAT 27600
QY 1571 -----CysProAlaGlyGly 1575
Db 27589 TTCCAGAGCCCTGGGGGTCCCTGAGTCCCTTCTCTCCGTTCTAGGCTGCCCCACCAAGAA 27540
QY 1575 sAsnValCysAsePLeuThrCysHisAlaArgGlyThrCysValAseGlnTTPArgPAl 1595
Db 27539 GAACCTGTGTGACACACACACTTGGCCATATATGTGTGGCACCTGTGTGAAACAGTGGGAGCC 27480
QY 1555 aPheSerCysGluCysProLeuGlyPheGlyGlySerCysValGln----- 1611
Db 27479 ATTCACTGCGCAGATGTCCTCTAGGCTTCCGGGGCAAGAGCTGTGCCAGGTAGAGTGG 27420
QY 1611 ----- 1611
Db 27419 CAGCTGCCACAGGTTGGGGCTGGGCTCTGTCAATGTCGTGGAAATTAAGGGAGGCTTGG 27360
QY 1611 ----- 1611
Db 27359 GGGCAGGCTCTGGGAGAGGCTCCGGAGATCAGGCTGTGGGTGGAAGAGCTGTCTGGGCA 27300
QY 1611 ----- 1611
Db 27299 GAGTAAGAGAGCTAAGCAGAGAGCAAAAGACTATGTCAGAGGAGGCCCTGGGACAGA 27240
QY 1611 ----- 1611
Db 27239 TGTAGGGCGGGGCTTGAAACAGGTAAATGCTCCGTGGGGGGGCAATAGCAGAGACCCGGT 27180
QY 1611 ----- 1611
Db 27179 GTAGAGGAGAGGAGAGGATCATCTCTCATAGAGCTGCCAGCTTCTGACTCCAGCCCT 27120
QY 1612 -----GlnMetAlaAsnProGlnHisPheLeuG 1621
Db 27119 GCCCTTGATCACTTTTTCATCTGCTTCCAGAAATGGCCAAATCCGACGCTTCTTGG 27060
QY 1621 ySerSerLeuValAlaTTPHisGlyLeuSerLeuProLeuSerGlnProTTPArgLeuS 1641
Db 27059 GAGAGAGCCCTTGTGAGCTGGCAATGGCTCTCTCTCCCACTCTCAAGCCCTGGACCTCA 27000
QY 1641 eTLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyA 1661
Db 26999 GCTCATGTTCGGACACAGCCAGGAGAGTGGCTGTCTGTGAGGCGGTGACACAGGGGC 26940
QY 1661 rGSetThrIleThrLeu----- 1666
Db 26939 GCAGACCATCACTCTGACAGTATGACAGGGCGGTGGAGTGTGGCAAGCCCGGCTTGG 26880
QY 1666 ----- 1666

```

---

```

Db 26879 GCTAATTGTAGAGCCTTCAGGGGGGTGTGAGCCCAAGGTGCTGGGCAAGGTTGGCAGA 26820
QY 1666 ----- 1666
Db 26819 AGCCAGACTTGGCTGGGCCCATTCGGTGGGTCAGTACTGCCACTTATTTGGCTTAAAG 26760
QY 1666 ----- 1666
Db 26759 AAATATCCCTCTCCTGCGCCCTCAGCGGAGACTGGAAGGCTTACTGAGTGTCT 26700
QY 1667 -----GlnLeuArgGlnGlyHisValMetLeuSerValG 1678
Db 26699 ACTAATTTTTTCTTCCCTCTCTGTGACACTTGGGAGGCGCAGGTAGTGTAAAGTGTGG 26640
QY 1678 lGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAlaAsnAspGlyA 1698
Db 26639 AGGCGACAGGGGCTCCAGGCTCATCTCTGCTGTGAGCCAGGCGGACCAATATGTGTG 26580
QY 1698 sPTTPHisHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuS 1718
Db 26579 ACTGGCATCAGCACAAGCTGTGCACTGGGAGCTAAGCGGGGGGCTTGGCCACGCCATCTGT 26520
QY 1718 ePheAsePThrGlyGlnGlnArgAlaGluGluValAseLeuGlyProArgLeuHisGlyLeuH 1738
Db 26519 CTTTGACTATGGGCAACAAGAGGAGAGGTATCTGGGGCTCTGCTGTGACGGGCTGCG 26460
QY 1738 lSLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaArgGlyP 1758
Db 26459 ACCTGAGCAATATACAGTGGGGGAGTCTCCGGGCCAGCCAGCATGTGGCCCGTGGCT 26400
QY 1758 hArgGlyCys----- 1761
Db 26399 TCCGGGGCTGTGTCAGGTGAATCTCTCCCGACCTTCATCCCAATTCCTACCTGGAG 26340
QY 1762 -----LeuGlnGlyValArgValSerAspT 1770
Db 26339 CCTGTTCCACTGAGGGCTAATGCTGCTTATATCCCTTCAGGGGTGAGGTAAAGCAGA 26280
QY 1770 hrProGluGlyValAsePLeuAsnProSerHisGlyGluSerIleAsnValGluGlnG 1790
Db 26279 CACCCAGAGGCTTATAGCAGTGTGATCCAGCCCGTGGGAGAGATCAATTTGAGAGCCAG 26220
QY 1790 lYSerSerLeuProAspProCysAsePLeuAsnProCysProAlaAsnSerThrCysSerA 1810
Db 26219 GCTGTAGCTGCGCAGATCCCTGTGATCTGATCATGATCCCAACAAGCTATCTGACGA 26160
QY 1810 sAspTTPAsePLeuThrCysSerCysSerCysAsp----- 1820
Db 26159 AGCATGGAGACTATTTCTGTAGCTGTGATCCAGGTAAAGTGAAGATGCTGGAAATG 26100
QY 1820 ----- 1820
Db 26099 GGGGTGGGCGCAGGGGTCAACATCCGTCCTTAATATAGCACTTAATCAGGGTATA 26040
QY 1820 ----- 1820
Db 26039 GCCATCTGTTGACAGGTTTCAAGCGGAATTAAGAGTGTATAGTGAAGAGATG 25980
QY 1821 -----ProGlyThrTrpGlyAspA 1827
Db 25979 CCAAGTAAGGGCGGGGCTTAACCTTTCCAGATCTTCTTGTGTCAGGTTACTATAGTGA 25920
QY 1827 snCysThrAsnValCysAsePLeuAsnProCysGlnHisGlnSerValCysThrArgLysP 1847
Db 25919 ACTGTAAATATGTGTGACTTGAACCAACAGGAGCAGCATGTGTGTGTATCCGGAABAC 25860
QY 1847 rOserAlaProHisGlyThrThrCysGluCysProProAsnThrYLeuGlyProThrCysG 1867
Db 25859 CCAATGACCCCAAGGCTTACATCTGCGAGTGTATCAAAATTAACCTTGGGCTTATGTG 25800
QY 1867 luthArgGly----- 1870
Db 25799 AGACCAAGTA-GGTGGGCGGGGCTTACTCTGTATATCACTCAGCCCGGAGAGTGAAGT 25741

```

QY 1870 ----- 1870  
 Db 25740 CGGGCTGAGCTGACGCCAATGAAAGACTCATGTGTTGAACCAAGCGGAGAGATGC 25681  
 QY 1871 ----- AspGlnProCysP 1875  
 Db 25680 AGTGTAGAGAGGGGCTCATGCTGTGTGGGTCTTTCTGACAGGATTTGACCACTTGCC 25621  
 QY 1875 roArgGlyTyrPTrpGlyHisProThrCysGlyProCysAsnGlyAspValSerIleGlyP 1895  
 Db 25620 CCGGTGCTGTGTGGAGACCCCAATGTGTCTGACCACTGCAAGGTGAGAGACCCGAATG 25561  
 QY 1895 heAspProAspCysAsnLysThrSerGlyGlyCysHisCysLys----- 1909  
 Db 25560 TTGAACCAAGCTGCACAAACAAACAGAGTGGCCACTGCAAGGTGAGAGACCCGAATG 25501  
 QY 1909 ----- 1909  
 Db 25500 AGCCCCGCTGTGTATACCTGTCTGTAGCTTGTGCCCCCTACAGGGGAGCAGAGAAT 25441  
 QY 1909 ----- 1909  
 Db 25440 GGGCTTCTCCAGAGAGCGAGGCAAGCATGTGTCTTTTGCTTATACAGGATGGAG 25381  
 QY 1909 ----- 1909  
 Db 25380 GGTGGGTGTTTGCCCTTATATCCGGCTGACGGGGATCCAGTTCAGTGTAGTCAGTGAG 25321  
 QY 1909 ----- 1909  
 Db 25320 CCTGCACTAGCTATGTGGGCTGTGTCAGCCCATTCAGTATTTGGCTTCTCACTGCT 25261  
 QY 1909 ----- 1909  
 Db 25260 ATGTCACTACCTGCTGTGAGAGTGTGAATCTCTCCCAAGAGACCCAGCATAGGGG 25201  
 QY 1909 ----- 1909  
 Db 25200 AGGCTCTCTGAGGAGGCGCACAGCTTGAGCTGTGCTAGCATCCAGAAAGACATCTT 25141  
 QY 1910 ----- GluAsnHisTyrAsnProProGlySerProT 1920  
 Db 25140 GATCCAGTGAACCCCTTCTTCCCTCTTAGAGATCACTACCAACCCCGACAGCCCCA 25081  
 QY 1920 hrcCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlnA 1940  
 Db 25080 CTGGCTCTGTGTGAGCTGTATACCACTGGCTCTTTGTCCGAGTGTGTGACCTGAGG 25021  
 QY 1940 spGlyGlnCysProCysLysPProGlyValIleGlyArgGlnCysAspArgCysAspAsnP 1960  
 Db 25020 ATGGCCAGTGTTCATGCAAGCTCGAGTCAATGGGGCTGAGTGTGATTCGTGACAAAC 24961  
 QY 1960 roPheAlaGluValThrThrAsnGlyCysGlu----- 1970  
 Db 24960 CTTTGTCTGAGTCAACCAATGGCTGTGAAAGTAGGGGGCTCTTAGCAGAGTAGGT 24901  
 QY 1970 ----- 1970  
 Db 24900 TCCCTCTCAACATATGCCAAGACTTACACTCCAGAGTTAAGAAATGAGGCTGGCATTG 24841  
 QY 1970 ----- 1970  
 Db 24840 GGTGCGGATGTGGAAGAAAGTTGCGGGCATGATAGAACTGCTTACCGATTAACCATG 24781  
 QY 1971 ----- ValAsnTyrAspSerCysProArgAlaIleGluAlaGlyTleTyrP 1986  
 Db 24780 TGTCTTCTCCATATGTAATTTAGACAGCTGCCACAGGCGCATAGAGGCTGGATCTGGT 24721  
 QY 1986 rPProArgThrArgPheGlyLeuProAlaAlaIaProCysProLysGlySerPhe----- 2004  
 Db 24720 GGGCCCGCAGCGGGTTGGGTACTCTGTGCCCCCTGCCCCAAGGGCTCTTTGGTA 24661

QY 2004 ----- 2004  
 Db 24660 GGTTTTAGATTACATATGATATTTGAACACTTGACCTTTACTCTGAGAGATAAAGATA 24601  
 QY 2004 ----- 2004  
 Db 24600 GCAGCTAGAGCCCAAGCTTGAAGCTCCGGGGCCCACTTCTGATTTATCTCCAGAG 24541  
 QY 2004 ----- 2004  
 Db 24540 ACTCAGTGTCTCACTTACCTTCACTCACTCACTCACTCACTCACTCACTCACTCC 24481  
 QY 2004 ----- 2004  
 Db 24480 TATATCAAGCCCAACAGAGCTGACGAAATCTGTGCTGTGAACAGGCTGGAGCT 24421  
 QY 2004 ----- 2004  
 Db 24420 AGACTGTGATTTGATGAGGGGTGCTTATCTGAGCTAACTACTTATGTACTGAGTTACC 24361  
 QY 2004 ----- 2004  
 Db 24360 CAACCTGAAGGAGGCCGTGCTGAGAGATAGGCGTGGTACTATCTCCAGAGATTG 24301  
 QY 2004 ----- 2004  
 Db 24300 AGATTAGTGTGACATGTTTCAATGGGAGGAGTGAAGTGAAGTGAAGGCACTAGTGA 24241  
 QY 2004 ----- 2004  
 Db 24240 TGAACAGGCGTAGTATATCTTGGGGCGGTGTGACAGAGTTACATAGCGTCCCTTGG 24181  
 QY 2004 ----- 2004  
 Db 24180 CCACGTACTGATTTTGTGAGAGATTTTGGCTGTGAGCTGTGACACTGGGCTTCTC 24121  
 QY 2005 ----- GlyThrAlaValArgHisCys 2011  
 Db 24120 TGTCTCCCTAAGCCCGGCTTCTTGGCTTCTTCTCCAGGAGACGTGTGGCCACTGT 24061  
 QY 2012 AspGlyHisIleArgGlyTyrLeuProProAsnLeuPheAsnGlyThrSerTlePheSer 2031  
 Db 24060 GATGAGCAGAGGGGCTGGCTCCGCAAACTTCACTGACGTCACTCACTTCTCA 24001  
 QY 2032 GluLeuLysGlyPheAla----- 2037  
 Db 24000 GAGCTTAAGGCTTTGT-AAAGTGAACCTGTCTTCTCTTCTTCAATCCCTGTCT 23942  
 QY 2037 ----- 2037  
 Db 23941 TTGAATAATGAGCTCCCGGGTGGGATTTAGCTCAGTGTGAGAGCGCTTGCCTAGAA 23882  
 QY 2037 ----- 2037  
 Db 23881 GCGCAAGCGCTGGGTGTGGTCCAGCTCCGGGAAAAAAGAACCAAAAAA 23822  
 QY 2037 ----- 2037  
 Db 23821 AAAAAGAAATAGAGCCCGAATCTGTCTTCTGAGGCTCTCTCACTGAGAGGCCAATTG 23762  
 QY 2037 ----- 2037  
 Db 23761 TTCCCACTGATCTTGTGAGGCTTTTACTGTGAGCCCAAGTCCCACTCTGTCT 23702  
 QY 2038 ----- GluArgLeuGlnArgAsnGly 2046  
 Db 23701 GACCTCTGCACTTCTGTGCTCTTCATCCAGGCGGAGCGCTTACAGGAAACAAATCAGGC 23642  
 QY 2047 LeuAspSerGlyArgSerGlnIleLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisThr 2066  
 Db 23641 CTGGACTCAGAGCGCTCCCAAGAGGCTGGCGCTCTGCGGTATGCCACAGACAGACC 23582  
 QY 2067 AlaGlyTyrPheGlySerAspValLysValAlaTyrGlnIleuAlaThrArgLeuLeuAla 2086

```

Db 23581 TCTGGCTACTTCGGGCGATGTCATGCAAGGTGGGCTTCACAGCTGGCCACAGACTGGCT 23522
Qy 2087 HIsGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValNHIsPheThrGlu 2106
Db 23521 CATGAGAGTGGCCAGCGGGGCTTTGGGCTGTCCGACACAGAGATGTGCACTTCCACTGAG 23462
Qy 2106 ----- 2106
Db 23461 GTGGGGCTTGAAGTAGGATGTGGGGCCCAAGTAGAGAGCCGTGTTCTGAGATCTTGG 23462
Qy 2106 ----- 2106
Db 23401 GGGAGATCTTGAAGGAGATACTTCTCTACTGTAAAGAGCTGCCACCCCTGGTACAGAG 23342
Qy 2106 ----- 2106
Db 23341 CCTCCATGATAGTCTGTAAACCAAGAGCCGACACAGATTGAAGTACCTTCTCTGG 23282
Qy 2106 ----- 2106
Db 23281 GAGTGGGATTTAGCTCAGTGTGAGAGCGCTTACCTAGAGAGCGAAGGCGCTGGGTTGG 23222
Qy 2106 ----- 2106
Db 23221 GTCCCAAGCTCCGAAAAAGAAAAAGAGTACTCTCTCTACAGTGTCAAG 23162
Qy 2106 ----- 2106
Db 23161 AATTAGTACTTACGTTGCACTTGAAGGCTGACCTGAGGCGAGCCCACTATTATACAA 23102
Qy 2106 ----- 2106
Db 23101 GAGTCAAGCTTGTGACGAGGACCAAGAGGCACAAAACAGAGCAGACATGCAAGAGAC 23042
Qy 2106 ----- 2106
Db 23041 ATGGGTGGCCCTGATTAAAGTGTGAAGAGGTTTGAAGCCAGTACCCATGCTGAA 22982
Qy 2106 ----- 2106
Db 22981 GTGACAGATTGATACACAGAGCTTCCAGGGGAGACGTGGGAGCTGTAAGGGAGAGA 22922
Qy 2106 ----- 2106
Db 22921 CCTAGTGGCGGGTCTGAACAGGCAAGGCTGAACCTTGATGAGTTGACCAAGC 22862
Qy 2107 ----- 2107
Db 22861 ATATGCCACTGTACCATGCCCCATCTCCAGAAATCTGTGAGGGTGGGAGCCCTCTCTG 22802
Qy 2117 AApThrAlaAsnLysArgHisPheGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 2136
Db 22801 GATGGGCGCATTAAGAGGCACTGGGACCTGATCCAGACAGAGAGGGTGGCACTGCTGG 22742
Qy 2137 LeuLeuGlnHisPheGlyAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisPheTyr 2156
Db 22741 CTGCTCCAGCACTATGAGGCTTATGCTGAGTGGCTGGCCAGAAATATCGGCACTTAC 22682
Qy 2157 LeuSerProPheThrIleValThrProAsnIle----- 2167
Db 22681 CTGAGCCCCCTTCACTATGTCACACCCAACTGTGTGAGGTGTGCTGGGCTGGGGGCG 22622
Qy 2167 ----- 2167
Db 22621 GGGGTGTGAGAGTCTGCGCATTTCAAAGGCGACGCGTGTGTGTGATGATCTTG 22562
Qy 2168 ----- 2168
Db 22561 CAACCATATTTCTTCTAGTCACTCTGTGGTGGCCCTGGATTAAGGGAACTTGTCTGG 22502
Qy 2182 AlaLysLeuProArgTyrGlnAlaLeuArgGlyGlnProProAspLeuGlnThrThr 2201

```

---

```

Db 22501 ACCAAGTCCCGCTTATGAGGCGTTACGTGGGAGCGCCTTCCAGATCTTGAAGCACA 22442
Qy 2202 ValIleLeuProGlnSerValPheArg----- 2210
Db 22441 GTCAATTTGGCCAGAGTCTGTCTTCAAGGTCAAGCGGGAGACGTGAGTGTAAACGGCTT 22382
Qy 2211 ----- 2211
Db 22381 GGGGTGGGCGACCCATGCAGTGGGGCCAGGCGCTGAGTGTAGAGCATTTACTTGCAGA 22322
Qy 2211 uThrProProValAlaArgProAlaGlyProGlyGlnAlaGlnLysProGlnLysLeuAl 2231
Db 22321 GATGCTCCCATGTGTGAGATCTGCAAGACCTGTGAAAGCCCAAGAGACTGAGAGCTGGC 22262
Qy 2231 AArgArgGlnArgArgHisPheProGlnLysSerGlnGlyAlaValAlaAspValIleI 2251
Db 22261 CCGGGGCGAGCGAGGCGACCCAGAACTGAGTCAGAGGCGAGGCGAGGCGCACTGTCACTAT 22202
Qy 2251 eTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProAspLysArgSerLeu-- 2270
Db 22201 TTACCAACACGCTGGCTGAGCTGCTGCCCACTACAGACCTAAGATTAAGCTTACCTGAG 22142
Qy 2270 ----- 2270
Db 22141 GTAAGCGCTACGACAAAGCTTAGAGGTTGGGGTTAGCAAGGAAATGTGAGCATCTC 22082
Qy 2271 ----- 2271
Db 22081 CCAAGTATAGGGGTGCTCTGCAAGCGCTTGTGTGACGTGTACTTCTGTATCAGAGT 22022
Qy 2272 IProLysArgProIleIleAsnThrProValAlaSerIleSerValHisAspAspGlnG 2292
Db 22021 CCCCAG--CGCCAGATCATCAACACACCTGTGTAGATCACTGTCCACAGATGATGAGA 21963
Qy 2292 uLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArgLeuGlnLysThr 2312
Db 21962 GCTCTTCCAGAGGCGACGTGACAGCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 21903
Qy 2312 uGlnArgThrLysProIleCysValPheTyrAsnHisSerIleLeu----- 2327
Db 21902 GGAGCGAACCAAGCCATCTGTCTTCTGGAACATTCATCTGTGAGCGTCCGTGA 21843
Qy 2327 ----- 2327
Db 21842 CCACCTTAAGCCCATGACTTTTGAACCAAGATCCAGGCTCTGCTGCTCACTTCTT 21783
Qy 2328 ----- 2328
Db 21782 CTTCCCTGACTCGCTCTTCTCTGCAAGGGTCAAGTGGCAAGGTGGCTGGTCCGCAAG 21723
Qy 2338 lYCyGlnValAlaPheArgAsnGlnLysSerHisValSerCyGlnCysAsnHisMetThr 2358
Db 21722 GCTGCGAGGTGTCTTCTTCCGTACAGAGCCATGTCACTGCTGCACTGCACTATGACA 21663
Qy 2358 ePheAlaValIleuMetAspValSerArgArgLys----- 2369
Db 21662 GCTTGTGCTGATGATGATGTGTCCGACAGAGGTGTGACTCCAGGGGTAGCCAGA 21603
Qy 2369 ----- 2369
Db 21602 GCAGCGGTGGGACCAAGTGCAGAGGCGCTCAGGCGCTCTTTTGTGACCCCAACAC 21543
Qy 2370 ----- 2370
Db 21542 TCATCCCTCAGAAATGGTGAATTTTGGCACTGAAGACCCCTGAGTATGTGGCCCTTGGAG 21483
Qy 2386 allThrLeuAlaAlaLeuLeuLeuThrPhePheLeuThrLeuLeuArgIleLeuArg 2406
Db 21482 TCACCTTAGTGGCTGATGATCACTTCTCTTCCCAACCTTCTGAGAGCCCTTGTGCT 21423
Qy 2406 eArgGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlyLeuValP 2426
Db 21422 CCAACCAAGATGGATCCAGGCACTTCAAGCTGCGCTGGGCTGGGCCCAAGCTGGTCT 21363

```

Qy	2426	heleuLeuGlyIleasnGlnAlaaspLeuPro-----	2436
Db	21362	TTCTCTGGGGATTCACAGAGCTGAACCT-CCCTGTAAAGATATTCTCTCCAGAGACCT	21304
Qy	2436	-----	2436
Db	21303	TCCCAACCTTCAGGCCACCAAGGCTGATGATGATGATCCCTTGTCCATCCCT	21244
Qy	2436	-----	2436
Db	21243	GCTCCCAAGCAATCAGTAGGTAAGGGCCCATTCAGTACCTGCCACAGTTGGCTT	21184
Qy	2436	-----	2436
Db	21183	CTCTCAGAACCAAGGCTTCCCACTCCCTGGGGGACCTCCTGTATGACGTCTTGG	21122
Qy	2437	-----PheAlaCyThrValIleAlaIleLeuLeuHisP	2448
Db	21123	CTCCGTATCATGTGCTGCGCCCTCAGTTTCTGTGACAGTCAATTGGCATCTGTGCACT	21064
Qy	2448	heleuTYrLeuCyThrPheSerTrpAlaLeuLeuGlnAlaLeuHisLeuTYrArgAla	2468
Db	21063	TCCGTACCTTTGACCTTCTCCGGGCTCTGCTGGAGGCTTACACCTGTACCGGGCCG	21004
Qy	2468	euthrGluValArgaspValAsnThrGlyProMetLeuArgPheTYrTYrMetLeuGlyTrpG	2488
Db	21003	TCACAGAGGGCCAGACGTCAATGCCAGGCCCATGCTTCTACTACATGCTGGGCTGGG	20944
Qy	2488	IyValProAlaPheIleThr-----	2494
Db	20943	GCGTTCTGTCTTCATCAGAGGATCTCGCTCCTCTTTCAGACCTAGGCTCCCTGTGG	20884
Qy	2494	-----	2494
Db	20883	CTAGGCTGTGGGTCCTACCTTCATGATGATACCCCAACATAGAAATTCCTGGAATCCC	20824
Qy	2494	-----	2494
Db	20823	TTCCCCCTAATCTTGGACAGCCTGGAATTAGGTGTCAAGGGCTAGGCTTCTCCTTGGAGAG	20764
Qy	2494	-----	2494
Db	20763	GGGAGTGAAGAGAGGTCCTTCTGTATGCTCCCACTGGGCTCTGTGAACCTTGATGACA	20704
Qy	2495	-----GlyLeuA	2497
Db	20703	TCATACTTGGCTCCACCCTTGAAGCACCTGTCTTGTGCTGTCTCTGTGCTCCAGGTCGTG	20644
Qy	2497	IaValGlyLeuAspProGlyGlyTYrGlyAsnProAspPheCySTrPLeuSerIleTYrA	2517
Db	20643	CTGGGGCTTGGATCCCGAAGGCTATGAGAAACCTTACTCTGTGCTCTCCATCTAAG	20584
Qy	2517	spThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSer-----	2532
Db	20583	ATAGGCTCATCTGAGTTTGTCTGAACAAGTGGCTTGTCTGTTCGTGTAAGTCTTGAA	20524
Qy	2532	-----	2532
Db	20523	AGCAATTTAGGTAAGCAGATGGCAGGGTCTACAGATTCCTCTTAAAGAAAGCTGAG	20464
Qy	2532	-----	2532
Db	20463	GCTCTGTGGGCCCCCTGGGCAAGAAAGAACTGTCTGTCTCTGTCTCTCATGTGCC	20404
Qy	2533	-----MetSerValPheLeuTYrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnA	2550
Db	20403	CACGTCAAGATAGGTGTCTTCTCTGTATATCTGTGAGCCGAGCCTCTGTGTGCCCAAC	20344
Qy	2550	rgGlnGlyPheGlnIlyAlaGlyProAla-----	2559
Db	20343	GTCAGGGCTTCGAGAAAGAAAGCCCTGTGTGTAGTACTGGCAGGTGCTGGGCAAGGGCG	20284

[illegible]



```

Db 19205 CTGAGGTGAGCCCGGAAATGGAGAGCTGGGAAAGAGGAAAGACTGCATGTAGAAC 19146
Qy 2684 ----- 2684
Db 19145 AAGCCGGCCAGCTGTAGAGAGCCAGTGTCTCAGACTGCCGCCAACAGAGTCAAGTT 19086
Qy 2684 ----- 2684
Db 19085 TAGAAGCATAGAGGACACAAAGTAGAGGCTGGGAATCTCATAAGAGGCCAGACTGAC 19026
Qy 2685 ----- 2685
Db 19025 CCTACAGCATAGTCTGTCTTTTAAAGGAGAGTCCAGCTGAACCTGGCCAGGTTCC 18966
Qy 2696 OProGlyLeuGlyAspProGlySerLeuPheLeuGlyGlyGlnAspGlnHis----- 2714
Db 18965 CCTGGGCTTAGGGGACCCCAAGTGGCTATTTCATGGAAGGCCAAGCCCAACACGAGTGA 18906
Qy 2714 ----- 2714
Db 18905 GACAGAGGCTTAGATGCTTAGAGGAGGAGGAGGCTGGCTTTTGTGAGGGCGGCTGG 18846
Qy 2715 ----- 2715
Db 18845 GGCAGGTGGGACAGTGTGATCCCTCCCTGACCTCCAGATCCGACACGAGACTCTGACA 18786
Qy 2722 eAspLeuSerLeuGlyAspAspGlnSerGlySerTyraIAserThrHisSerSerAsp 2742
Db 18785 GTGACCTGTCCCTGGAGATACCAAGATGCTTCTACGCTTCAACCACTCATCAGACA 18726
Qy 2742 eArgLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2762
Db 18725 GC-----GAGGAAAGAGGAGGAGGAGGCGCCCTTCCCTGGCAGAGGAGGCTGGG 18678
Qy 2762 sPserLeuLeuGlyProGlyAlaGlyLeuProLeuHisSerThrProlys----- 2779
Db 18677 ACAGCTGTGTGGGCTCTGAGAGTGAAGACTGCCCTTCAAGTACCCCA- GAGTGGG 18619
Qy 2779 ----- 2779
Db 18618 CACAGCACAGGAGCTGAGCCATTGGGGGGGCAAGGAGCTAAGAGGCTTCTT 18559
Qy 2780 ----- 2780
Db 18558 GGCATCTCTGTCTACCTCTCTGGAGCTCATAGGCTTCTTTCTTTGACAGATGGGGTTC 18499
Qy 2783 rGlyProGlyTyraIaProIleProGlyAspPheGlyThrThrAlaLysGlnSerSerg 2803
Db 18498 CAGGCTCTGGAGAGTCCCTTGGCCAGAGACTTTGGGACCAACAAAGAGATAGTG 18439
Qy 2803 TyAsnGlyAlaProGlyLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2823
Db 18438 GTAGTGGGCCCCCTTGAAGAGGGGCGGAGGAGATGAGATGCTTAACTCGGAAAGGT 18379
Qy 2823 eLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyLe----- 2838
Db 18378 CTCTGGGACCCCTTCCGGGCTTCTTACCAACTCAAAAGGTTA- GTGAGAGTCTCC 18320
Qy 2838 ----- 2838
Db 18319 TCCATCTGCTTCACTCTGCGATCTCTTGCTGGAATTTAGAAATGCTCATCTCANGTA 18260
Qy 2839 ----- 2839
Db 18259 CCCACACCTCTTACCCAGGATCTCAAGAGAAAGTGTCTGCCACCATCAGGAAAGA 18200
Qy 2851 eSerLeuLeuAlaGlyLeuProLeuGlnGlySerThrGlySerSerAspGlySerSerAla 2871
Db 18199 GTAGCTCTTAAGGTTCCTCTGAGAGGAGCAAGGAGTCTTCCCGGGGCTTCCACCGCCA 18140
Qy 2871 eArgLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLugLug 2891

```

```

Db 18139 GTGAGGACAGAGAACGAGACTTCTCCAGCCCAACCGCGCCAGTCTTCCAGGAAC 18080
Qy 2891 LLeuAsnGlyValIleProIleAlaIleSerIleTyraIaGlyThrValAspGlnAsp 2911
Db 18079 AGCTGAATGGGGTCAATGCCATGCAATGAGCATCAAGGAGCACTGTGATAGGACT 18020
Qy 2911 eSerGlySerGln 2915
Db 18019 CTCTGGGCTCCAG 18006

```

Search completed: February 14, 2004, 03:42:07  
 Job time : 26841 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 17:42:21 ; Search time 1423 Seconds

(without alignments)  
5544.945 Million cell updates/sec

Title: US-09-916-849A-3  
Sequence: 1 MMSPTATGVPPLTPPPPLLL.....AGTVDESSGSEFLFPNPLH 2923

Scoring table: BLOSUM62  
Xgap 10.0 , Xgapext 0.5  
Ygap 10.0 , Ygapext 0.5  
Delop 6.0 , Deloxt 7.0  
Delop 6.0 , Deloxt 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q/cq2 1 USPTO.spool1/US09916849.rng -MINMATCH=0.1 -LOOPL=0  
-DB=N Geneseq 19Jun03 -QPM=fastap -SUFFIX=trng -MINLEN=0 -MAXLEN=200000000  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human0.cdi  
-LIST=45 -DOCALL=200 -THR\_SCORE=pc -THR\_MAX=100 -THR\_MIN=0 -ALLEN=15  
-MODE=LOCAL -OUTWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849.OCN 1 1 0 -runat 11022004.143815.20929 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELOXT=7

Database :

N\_Geneseq 19Jun03:.\*  
1: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1980.DAT:.\*  
2: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1981.DAT:.\*  
3: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1982.DAT:.\*  
4: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1983.DAT:.\*  
5: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1984.DAT:.\*  
6: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1985.DAT:.\*  
7: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1986.DAT:.\*  
8: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1987.DAT:.\*  
9: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1988.DAT:.\*  
10: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1989.DAT:.\*  
11: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1990.DAT:.\*  
12: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1991.DAT:.\*  
13: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1992.DAT:.\*  
14: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1993.DAT:.\*  
15: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1994.DAT:.\*  
16: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1995.DAT:.\*  
17: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1996.DAT:.\*  
18: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1997.DAT:.\*  
19: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1998.DAT:.\*  
20: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA1999.DAT:.\*  
21: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2000.DAT:.\*  
22: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001A.DAT:.\*  
23: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2001B.DAT:.\*  
24: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2002.DAT:.\*  
25: /SIDS1/gcgdata/geneseg/geneseqn-emb1/NA2003.DAT:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	15545	100.0	8772	22	AA511678	Human Flamingo cDN
2	15545	100.0	10531	25	AB242868	Human GPCR CELSR2
3	15518.5	99.8	8871	22	AA511677	Human Flamingo cDN
4	15393	99.0	11762	25	ABX34546	Human mdr1 cDNA SE
5	15279	98.3	9321	24	ABX15177	Human REPR 9 cDNA
6	13178.5	86.1	9401	22	ABA08648	Human FLAMINGO 1 h
7	12839.5	82.6	9121	21	AACT6401	Human ORFX ORF1956
8	8781.5	56.5	9045	22	AA506332	DNA encoding seven
9	8781.5	56.5	11389	25	AA559825	Human novel cytoxi
10	8781.5	56.5	11389	25	AB242827	Human GPCR CELSR1/
11	8736.5	56.2	9087	22	AA514986	Human NOV7 DNA. H
12	8710	56.0	10195	18	AA185320	Mouse receptor ME2
13	7836.5	50.4	11965	25	AB242586	Human CELSR3 nucle
14	7815.5	50.3	12348	24	ABO82327	Human NOV7 encodin
15	7766	50.0	11648	22	ABD08065	Human extracellular
16	7760.5	49.9	13935	23	AA574583	DNA encoding novel
17	4823	31.0	12791	23	AB11557	Drosophila melanog
18	4766.5	30.7	17282	23	AB11556	Drosophila melanog
19	4584.5	29.5	6791	18	AA185319	Mouse receptor ME2
20	3958	25.5	2332	22	AB19447	Human nervous syst
21	3958	25.5	2332	22	AAK83060	Human immune/haema
22	3958	25.5	2332	22	AAK83060	Human immune/haema
23	3958	25.5	2332	22	ABQ66814	Human DNA for a no
24	3818	24.6	2391	25	AAV07219	Human breast cance
25	3787	24.4	2603	19	AAV07219	Human calcitonin r
26	3769.5	24.2	3912	22	AAV98728	Human late stage o
27	3734	24.0	4152	21	AAV15924	Human polynucleoti
28	3701	23.8	2695	24	AB190074	Human secretory po
29	3526	22.7	2077	22	AB19448	Human nervous syst
30	3526	22.7	2077	22	AAK83061	Human immune/haema
31	3526	22.7	2077	22	AA531491	Human DNA for a no
32	3526	22.7	2077	22	ABQ66815	Human polynucleoti
33	2838	18.3	1734	22	AB19445	Human nervous syst
34	2425.5	15.6	4874	22	ABA09134	Human protocadheri
35	2078	13.4	3620	25	AB236319	Human secretory po
36	1733.5	11.2	15603	25	ABX56292	Human NOV17a CG928
37	1567.5	10.1	15441	23	AB106029	Drosophila melanog
38	1520.5	9.8	13960	23	AA587164	DNA encoding novel
39	1520.5	9.8	14756	24	ABV94753	Human pancreatic c
40	1520.5	9.8	14756	25	AA54469	Human cadherin (CA
41	1520.5	9.8	20706	23	AB106028	Drosophila melanog
42	1511.5	9.7	14785	23	ABQ60905	mRNA for hFet prot
43	1477.5	9.5	14705	23	AB12585	Drosophila melanog
44	1451	9.3	925	22	AAK62671	Human immune/haema
45	1451	9.3	925	22	AA531256	Human cDNA encodin

#### ALIGNMENTS

RESULT 1  
AA511678 standard; cDNA; 8772 BP.  
AA511678;  
24-OCT-2001 (first entry)  
Human Flamingo cDNA splice variant.  
Human Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
signal transduction pathway; Bacterial; Fungal; viral; protozoan; cancer;  
anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
obesity; hypotension; hypertension; urinary retention; angina pectoris;  
myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
benign prostatic hypertrophy; psychotic disorder; neurological disorder;  
manic depression; delirium; dementia; severe mental retardation; se;



QY 441 G1yAsnAlaArgG1yGlnPheTyrLeuAspAlaGlnThrG1yAlaLeuAspValIleSer 460  
Db 1321 GGCAATGCTGGGGAGAGTTTATCTGGATGCCAGATCGAGCTCGAGTGTGTAGC 1380  
QY 461 ProLeuAspTyrG1uThrThrLysG1uTyrThrLeuArgValAlaG1aGlnAspG1yG1y 480  
Db 1381 CCTTTTACATAAGACACCAAGAGATGACCTTACGGGTGGACAGACAGATGTGGC 1440  
QY 481 ArgProLeuSerAsnValIleSerG1yLeuValThrValGlnValLeuAspIleAsnAsp 500  
Db 1441 GGTCCCCACCTCTTAATGTCCTGGCTGGTGAAGTACAGTCTCGAGATATCAAGAC 1500  
QY 501 AsnAlaProIlePheValIleSerThrProPheGlnAlaThrValLeuG1uSerValProLeu 520  
Db 1501 AATCCCCCACTTCGTACAGACCCCTTTCCAGGCTACTGCTCGAGAGAGGTCCCTTA 1560  
QY 521 G1yTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaG1yAspAsnAlaArgLeu 540  
Db 1561 GGTACCTGTGTCTTCATGTCCAGCTATCCAGCTGCTGTGCAATGTCCCGCTG 1620  
QY 541 GluTyrArgLeuAlaG1yValG1yHisAspPheProPheThrIleAsnAsnG1yThrG1y 560  
Db 1621 GAATACGCTTGTGTGGGTGGAGATGACTTCCCTTCACATCAACAATGACAGGC 1680  
QY 561 TrpIleSerValAlaAlaG1uLeuAspArgG1uValAspPheTyrSerPheG1yVal 580  
Db 1681 TGGATCTCTGTGGCTGCACTGACCGGGAGAAAGTATTCTACAGCTTTGGGTA 1740  
QY 581 G1uAlaArgAspHisG1yThrProAlaLeuThrAlaSerAlaSerValIleThrVal 600  
Db 1741 GAAGCTGAGACCAATGACATCCAGCACTCATCTCGCGCAAGTGTCAAGGTGATCGTC 1800  
QY 601 LeuAspValIleAsnAsnAsnProThrPheThrGlnProG1uTyrThrValArgLeuAsn 620  
Db 1801 CTGGATTCACAGACACATTCACCTTTACCAACAGATACACAGTGGCTCAT 1860  
QY 621 GluAspAlaAlaValG1yThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
Db 1861 GAGGATCAGCTGTGGGACCAAGGTGTGACGTGTGACGTGTGACCGTGAATGCTCAT 1920  
QY 641 SerValIleThrTyrGlnIleThrSerG1yAsnThrArgAsnArgPheSerIleThrSer 660  
Db 1921 AGGTCACTACCTTACAGATCAACAGTGGCAATCTGAAACCCCTCTCATCAACAGC 1980  
QY 661 GlnSerG1yG1yLeuValIleSerLeuAlaLeuProLeuAspTyrLysLeuG1uArgGln 680  
Db 1981 CAAGTGTGTGGGTGGTGTGTCTTGTCCCTGCTGCACTGACATGAACTTGAAGCGGAC 2040  
QY 681 TyrValLeuAlaValIleThrAlaSerAspG1yThrArgGlnAspThrAlaGlnIleVal 700  
Db 2041 TATGTGTGGCTGTACCGCTCCGATGGACCTGGAGGACAGGACAGATGTGTG 2100  
QY 701 AsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
Db 2101 AATGTACCGACCGCAACCAACGATGCTGTCTTTCAGAGCTCCCACTTACAGTGAAT 2160  
QY 721 ValAsnG1uAspArgProAlaG1yThrThrValValLeuIleSerAlaThrAspG1uAsp 740  
Db 2161 GTTATATAGACCGGCGGCGGACCAAGGTGTGTGATCAAGCGGCAAGATGAGAC 2220  
QY 741 ThrG1yG1uAsnAlaArgIleThrTyrPheMetG1uAspSerIleProGlnPheArgIle 760  
Db 2221 ACAGGTAGAAATGCCCATCACTTCACTTCAATGAGAGACAGATCCCGAGTTCCGATC 2280  
QY 761 AspAlaAspThrArgValAlaValThrThrGlnAlaG1uLeuAspTyrG1uAspGlnValSer 780  
Db 2281 GATTCACACCGGGGCTGTACCAACCGAGCTGAGCTGACGTAAGACCAAGTGTCT 2340  
QY 781 TyrThrLeuAlaIleThrAlaArgAspAsnG1yIleProGlnLysSerAspThrThrTyr 800  
Db 2341 TACACCTGGGCACTTATCTGTGGGACATGTGCAATTCGCCAGAGTCCGACCAACCTTAC 2400

QY 801 LeuG1uIleLeuValIleAspValIleAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db 2401 CTGGAGATCTCTGGTGAACGACGTGAATGACAAAGCCCTCAAGTCTCCGACAGACTCTAC 2460  
QY 821 GlnG1ySerValIyrg1uAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2461 CAGGGCAGTGTCTATAGAGATGTGCCACCTTCACTTACGCTCTGACAGATCAAGCCACT 2520  
QY 841 AspArgAspSerG1yLeuAsnG1yArgValPheTyrThrPheGlnG1yG1yAspAspG1y 860  
Db 2521 GATGTGATTTCTGACTTAATGACAGGTCTTCTTCACTTCCAGGAGGCGACGATGGA 2580  
QY 861 AspG1yAspPheIleValG1uSerThrSerG1yIleValArgThrLeuArgValLeuAsp 880  
Db 2581 GACGGTACCTTATTTGTATGTGATCCAGTCAAGCATGTGTGACACCTTACGAGGCTGGAT 2640  
QY 881 ArgG1uAsnValAlaGlnTyrValIleuAspAlaTyrAlaAspLysG1yMetProPro 900  
Db 2641 CGAGAGACGTGGCCCAATATGTCTTGGCGGCATATGCAGTGACAAAGGGATGCCCA 2700  
QY 901 AlaArgThrProMetG1uValThrValIleuAspValIleAsnAsnProProVal 920  
Db 2701 GCCGCAACCTATGAGAGTGAAGTACATCTGTGTGATGTGAATATCAATCCCTGTCTC 2760  
QY 921 PheG1uG1uAspG1uPheAspValPheValG1uG1uAsnSerProIleG1yLeuAlaVal 940  
Db 2761 TTTGACAGGATAGTGTATATGTGTGTGTGAGAGAGAACAGCCCATTTGGGCTAGCCGTG 2820  
QY 941 AlaArgValIleThrAlaThrAspProAspG1uG1yThrAsnAlaGlnIleMetTyrGlnIle 960  
Db 2821 GCCCGGTCAACGCCCTGACCTGACCCGATGAAAGGACCAATGCCAGATTAATGACAGATT 2880  
QY 961 ValG1uG1yAsnIleProG1uValPheGlnLeuAspIlePheSerG1yG1uLeuThrAla 980  
Db 2881 GTGAGGGACATATCTTGAAGTCTTCCAGCTGGAATCTTCTCCGGAGAGCTGACAGCC 2940  
QY 981 LeuValAspLeuAspTyrG1uAspArgProG1uTyrValIleuValIleGlnAlaThrSer 1000  
Db 2941 CTGTGACTTATAGACTTACAGAGACCGGCTGTAGTACGTCTGTGTATCCAGGCAAGTCA 3000  
QY 1001 AlaProLeuValIleThrArgAlaThrValHisValArgLeuLeuAspArgAsnAsnPro 1020  
Db 3001 GCTCTCTGTGTAGCGGGCTACAGTCAAGTCCGCTCTCTTGAACCGCAATGCAACCCA 3060  
QY 1021 ProValIleuG1yAsnPheG1uIleuPheAsnAspTyrValThrAsnArgSerSerSer 1040  
Db 3061 CCAAGTCTGGGCAACTTTAGATCTTTTCACACATTAATGTCAACATGTCTCAAGCAGC 3120  
QY 1041 PheProG1yG1yAlaIleG1yArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
Db 3121 TTCCTGTGGGGTGCATTTGGCGAGTACCTGGCCATGACCTGATATCTCAATAGTCTG 3180  
QY 1061 ThrTyrSerPheG1uArgG1yAsnG1uLeuSerLeuValIleuLeuAsnAlaSerThrG1y 1080  
Db 3181 ACTTACAGCTTGTGAGGGGAAATGAACTCAAGCTGTGTCTGTCAATCTTCAAGGGGT 3240  
QY 1081 GluLeuIleLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSerVal 1100  
Db 3241 GAGCTAGACTTAAGCCGCGCACTGACACACACACCCGCTCTGAGGCACTATGAGCGTG 3300  
QY 1101 LeuValIleSerAspG1yValHisSerValIleThrAlaGlnCysAlaLeuArgValThrIle 1120  
Db 3301 CTGTGTACAGACGGCTGACACAGCGTGAACCGCCAGTGGCGGTGTGTGACATCATC 3360  
QY 1121 ThrAspG1uMetLeuThrHisSerIleThrIleuArgLeuG1uAspMetSerProG1uArg 1140  
Db 3361 ACCGATGAGATGTCAACCAAGACATCAAGCTGCGCTGTGAAGACATGTCAACCGAGCC 3420  
QY 1141 PheLeuSerProLeuLeuG1yLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
Db 3421 TTCCTGTCAACATGTAGGCTCTTCACTCAAGCGGTGTGCCCAACGCTGTGGCAAGCCA 3480  
QY 1161 ProAspHisValIleValIlePheAsnValGlnArgAspThrAspAlaProG1yG1yHisIle 1180

```
Db 3481 |||||CGGAGCCAGTGTGCTTCAACGTACAGCGGAGCACCGAGCCCGCGGGGCGCACTC||| 3540
Qy 1181 LeuAenValSerLeuSerValGlyInpProProGlyProGlyValGlyProProAenLeu 1200
Db 3541 CTCAACGTAGAGCTGTGGTGGCCAGCGCCAGGAGCCCGGGGGGGCGCCCTTCCTG 3600
Qy 1201 ProSerGluAapLeuGlnGluArgLeuValLeuAenAArgSerLeuThrAlaIleSer 1220
Db 3601 CCTCTGAGAGACTTCAGAGAGGCGCTATACCTCAACCGCAGCTCTCTAGCCGCACTTCG 3660
Qy 1221 AlaGlnAArgValLeuProAenAapAenIleCysLeuAArgLupProCysGluAenIle 1240
Db 3661 GCACAGCCCGTGTCTCCCTTCAGACCAACATCTCGCGGAGCCCTCGCAGAACCTAC 3720
Qy 1241 MetAArgCysValSerValLeuAenAArgAapSerSerAlaProPhaIAlaIleSerSer 1260
Db 3721 ATGCGCTGTGTGCTGAGTGTGCTGCTTCAGCTCTCGCGCCCTTCATGCGCTCTCTCC 3780
Qy 1261 ValLeuAenAArgProIleHisAProValGlyValLeuAArgCysAArgCysAProProIlyPh 1280
Db 3781 GTGCTCTTCGGCCCATCAACCCGTCGAGGGGCTGGCTGCGCTGCGCCCGCGCTTC 3840
Qy 1281 ThrGlyAapIlyCysGluThrGlyValAapLeuCysIlyTserAArgProCysGlyProHis 1300
Db 3841 ACGGTGTACTGTGCGAGACCGAGGTGTGACTCTGTCTACTCGCGCCCTGTGGCCCGCAC 3900
Qy 1301 GlyAArgCysAArgSerAArgGluGlyValIlyThrCysLeuAArgAArgIlyThrGly 1320
Db 3901 GGGCGCTCCCGAGCCCGGAGGGCGGCTACACTCTCTGTGTGAGTGGCTACAGGGGT 3960
Qy 1321 GluHisCysGluValSerAlaArgSerGlyAArgCysThrProGlyValCysLeuAenGly 1340
Db 3961 GAGCACTGTGAGGTGAGTGTGCTGCTCAAGCCGTTTCACCCCGGGTGTGCAAGATGG 4020
Qy 1341 GlyThrCysValAenLeuLeuValGlyValPhelyCysAapCysProSerGlyAapPh 1360
Db 4021 GGCACCTGTGTCAACTGTGTGGCGGTTTCAGATGCAATGGCCATCTGTGAGACTTC 4080
Qy 1361 GluLysProIlyCysGlnValIlyThrAArgSerPheProAlaHisSerPheIleThrPhe 1380
Db 4081 GAGAAAGCCCTACTGCTCAAGTGTACACGCGAGCTTCCCGCCCACTCTTATCACTTT 4140
Qy 1381 ArgGlyLeuAArgGlnAArgPheHisPheThrLeuAlaLeuSerPheAlaThrIlyGluArg 1400
Db 4141 CGCGGCTCGCCGAGCGTTTCACTTCACTTGGCCCTCTCGTTTGCACAAAGAGAGGC 4200
Qy 1401 AapGlyLeuLeuLeuIlyAenGlyAArgPheAenGlyValHisAapPheValAlaLeuGlu 1420
Db 4201 GACGGGTGTGTGTATCAATGGCGTTTCATATGAAAGACATGACTTGTGGCCCTCGAG 4260
Qy 1421 ValIleGlnGlnValGlnLeuThrPheSerAlaGlyIlySerThrThrThrValSer 1440
Db 4261 GTGATCCAGAGAGAGTCAAGTCACTTCTCTGAGGGAGTCAACACACGAGTGC 4320
Qy 1441 ProPheValIProGlyValIleSerAapGlyIlyIlyIlyIlyValGlnLeuIlyTyr 1460
Db 4321 CCATTCGTGCGCCGAGGAGTCAAGTGTGCGAGTGTGCACTGCACTGAAATATCTAC 4380
Qy 1461 AenLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnIlyValAla 1480
Db 4381 AATAAGCCACGTGTGTGTACAGACAGGCTCCACAGGGCCCATAGAACAAAGTGTCT 4440
Qy 1481 ValValThrValAapGlyCysAapThrGlyValAlaIleAuaArgPheGlySerValLeuGly 1500
Db 4441 GTGGGTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4500
Qy 1501 AenIlySerCysAlaIleGlnGlyThrGlnGlyIlySerIlyIlySerLeuAapLeuThr 1520
Db 4501 AACTACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4560
Qy 1521 GlyProLeuLeuLeuGlyValIlyProAapLeuProGlyIlySerPheProValArgMetAArg 1540
|||
```

---

```
Db 4561 GGGCCCTGCTACTAGAGCGGGGTGCTGACCTGCGCAGAGACTTCCAGTCCGAATGCGG 4620
Qy 1541 GlnPheValGlyCysMetAArgAenLeuGlnValAapSerAArgHisIleAapMetAlaAap 1560
Db 4621 CAGTTCGTGGCTGTCAAGCGGAACCTGCAAGTGTGACAGCGGACATATGACTGTGTGAC 4680
Qy 1561 PheIleAlaAenAenGlyThrValIProGlyCysProAlaIlyIlyAenValCysAapSer 1580
Db 4681 TTCATTGCCAACAATGGCACTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 4740
Qy 1581 AenThrCysHisAenGlyIlyThrCysValAenGlnIlyAapAlaPheSerCysGlyCys 1600
Db 4741 AACTGTGCCAATATGGGGCACTTGTGTGAACCAAGTGTGACCGGTTCACTGTGAGTGC 4800
Qy 1601 ProLeuGlyPheGlyIlyIlySerCysAlaGlnGlyMetAlaAenProGlnHisPheLeu 1620
Db 4801 CCTTGGGCTTGTGGGGCAAGCTGTGCGCCAGGAATGTGCATTCACAGCACTTCTTG 4860
Qy 1621 GlySerSerLeuValAlaIlyThrHisGlyLeuSerLeuProIleSerGlnProIlyTyrLeu 1640
Db 4861 GCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 1641 SerLeuMetPheAArgThrAArgGlnAapGlyValLeuLeuGlnAlaIleThrAArgGly 1660
Db 4921 AGCTCATGTTCGCGACGCGCCAGGCGCAGAGTGTCTGTGTGAGGCTATCACAGGGGG 4980
Qy 1661 ArgSerThrIleThrLeuGlnLeuAArgGluGlyHisValMetLeuSerValGluGlyThr 1680
Db 4981 GCAGACACATCACCTTACAGTGTACAGAGGGGCGCATGTATGTGTGAGCGTGTGAGGGCA 5040
Qy 1681 GlyLeuGlnAlaSerSerLeuAArgLeuGluProGlyAArgAlaAenAapGlyAapIlyHis 1700
Db 5041 GGGCTTCAAGCTCTCTCTCTCCGTGTGAGCCAGGCGCGGCGCAATGACGTGTGCTGGCAC 5100
Qy 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyIlyProGlyHisAlaIleLeuSerPheAap 1720
Db 5101 CATGTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5160
Qy 1721 TyrGlyGlnGlnAArgAlaGluGlyAenLeuGlyProAArgLeuHisIleGlyLeuHisIle 1740
Db 5161 TATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5220
Qy 1741 AenIleThrValGlyIlyIleProGlyProAlaGlyValAlaAargGlyPheAArgGly 1760
Db 5221 AACTATACATGTGGCGGATATCTGTGGCCAGCGCGGTGTGCGCTTTCGGGGGC 5280
Qy 1761 CysLeuGlnGlyValAArgValSerAapThrProGluGlyValAenSerLeuAapProSer 1780
Db 5281 TGTGTGAGGGGTGTGCGGTGTGAGCGATACGCGAAGGGGATTAACAGCTGTGATCCAGC 5340
Qy 1781 HisGlyIlySerIleAenValGlnGlnGlyCysSerLeuProAapProCysAapSerAen 1800
Db 5341 CATGGGAGAGACATCAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5400
Qy 1801 ProCysProAlaAenSerIlyCysSerAenAapIlyAapSerIlySerCysAap 1820
Db 5401 CCGTGTCTCTCTAACAAGTATGTGACAGACAGTGTGAGCAGTATCTCTCACTGTGAT 5460
Qy 1821 ProGlyIlyTyrGlyIlyAapAenCysThrAenValCysAapLeuAapProCysGlnIleGln 1840
Db 5461 CAGGTACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5520
Qy 1841 SerValCysThrAArgIlyProSerAlaProHisGlyTyrThrCysGluCysProProAen 1860
Db 5521 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5580
Qy 1861 TyrLeuGlyProIlyCysGluThrAArgIleAapGlnProCysAapProAArgIlyTyrTyrGly 1880
Db 5581 TACCTTGGGCAATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5640
Qy 1881 HisProThrCysGlyIlyProCysAenCysAapValSerIlyGlyPheAapProAapCysAen 1900
Db 5641 CATCCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5700
```

QY	1901	LYSTHSGRTGIGLVCYEHISCYALVSGIUBANHSITATGPRPROGIIYSEITROTH	1920
Db	5701	AAGCAAGCGCGGAGTGGCACTGCAGAGAGAAACCACTACCGGCCCGCCAGGACGCCCAAC	5760
QY	1921	CYSLEULEUCYSAAPCYETTPROTHGLYSEILEUSERRANGVALCYSAAPPROGLIUNAP	1940
Db	5761	TGCTCTGTGTGAGTGTCTACCCCAAGACAGCTCTCTGTCCAAGTGTGTGACCTCGAGAT	5820
QY	1941	GLYGLNCYPROCYSLYPROGLIYVALILEGLIARGGLNCYASPARCYASPARANPRO	1960
Db	5821	GGCCAGTGTTCATGCAAGCAAGGTGTCTATCGGGGCTCAGTGTGACCGGTGTGACAACCT	5880
QY	1961	PHENALAGIUNVALTHRTTHASNGIYCSGLUNVALANITYAPSERCYEPROALALILE	1980
Db	5881	TTTGTGTAGGTGACCAACCAATGGCTGTGTAAAGTAATTTGACAGCTGCCCAAGACGAT	5940
QY	1981	GIUNALAGIYILETRTPPROAROTHRAPRPNGLIYLEUPROALALALAPROCYAPRO	2000
Db	5941	GAGCTGGGATCTGTGTGGCCCCGTACCCGCTTCGGAGCTGCTGCTGTCTCTGTCTCC	6000
QY	2001	LYSGIYSEPRHEGIYTHRALVALARGHISCYASPARGLIUNISARGGLIYTPLEUPRO	2020
Db	6001	AAAGGCTCTTTGGGACTGTGTGGCCAGCATGTATGACACAGGGGGTGGCTCCCCCA	6060
QY	2021	ASNULEPHEANCYETHSERILETHRESERGLIUNGLYGLYPHENALGLIUNARGLEU	2040
Db	6061	AACTCTTTCACCTGCACGTTCACACTCTTCACAACTGAAGAGGGCTTGCTGAGGGGTGA	6120
QY	2041	GLINARGAENGILUSEGLIYLEUAPSERGLIYRGSERGLIUNLEUALILEULEUNARG	2060
Db	6121	CAGCGGATGATGCTCAGGCTTACACTCAGGGGCGCTCCAGACACTAGCCCTCTCTGGC	6180
QY	2061	ASNALATHRGILNHISTHRIAGIYTPRHEGLIYSEAPVALIYSEVALIATYGLIUNLEU	2080
Db	6181	AAACCCACGACGACACACAGCTGTGCTACTTCGGACAGCGACGTCAAGGTGGCTTACAGCTG	6240
QY	2081	ALATHRAAGLEULANALAHISGLIUSERETHRGILNARGGLIYPHEGLIYLEUSERIALATHRGIL	2100
Db	6241	GCCAGCGGGCTGTGGCCCAAGAGAGACCCAGCGGGGCTTTGGGCTGTGTGGCAACAG	6300
QY	2101	ASAPVALHISPHETHRGILUANASNULEUNARGVALGISERIALALEULEUNAPTHRALAHAN	2120
Db	6301	GACGTGACCTTCACGTGAATCTGTGTGGGGTGGGCAAGGGCTCTCTGACACAGCCAC	6360
QY	2121	LYEARGHISTPGLIUNLEULEGLIUNGILNTHRGILGLIYTHRALATPPLLEUNGILNHIS	2140
Db	6361	AAAGGCGCACTGGAGCTGATCAGAGAGACAGAGGGTGGCACCGCTGGCTCTCCAGAC	6420
QY	2141	TYTRGLIUNALATYRALASERIALALEUNAGILNANMETHARGHISTHRYTLEUSERPROTHE	2160
Db	6421	TATAGAGCTTAGGCCAGATGAGCCCTGGCCCAAGACATGGGCAACCTTACCTTAAGCCCCCTTC	6480
QY	2161	THRIILEVALTHRTPROASNHILEVALIIESEERVALIARGLEUNAPLYSGIYASNPHEAL	2180
Db	6481	ACCATCTGTCAGGCCCAATTTGTATCTCCGTATGTGGCTTGGACAAAGGGAACCTTGCT	6540
QY	2181	GLIYALYLYLEUPROARGTYTRGLIUNALEUNARGGLIYGLIUNGILNPROPROASPLEUGILNTH	2200
Db	6541	GGGGCCAGACGTGCCCCGTCTAGAGGCCCCGTGTGGGAGACGCCCCCGGACCTTGAGACGA	6600
QY	2201	THRTVALIIELEUPROGLIUSERVALPHEARGILNTHRTPROPOVALIVALARGPROALAGIY	2220
Db	6601	ACAETCATTTGCTGCTGAGTCTGTCTTCCAGAGAGAGCCCCCCCCGTGTGTAGGCCCGGACGC	6660
QY	2221	PROGLIYGLIUNAGIUNGILNPROGLIUNLEUNALARGARGGLIUNARGHISPROGLIUNLEU	2240
Db	6661	CCCCGAGAGGCGCCAGAGACGAGAGAGCTGGCACGGCGACAGCCAGCGCAACCCGAGAGCTG	6720
QY	2241	SERGLIUNGILYUNALVALALASERVALIILETETRYRATGTHRTLEUNALAGIYLEUNLEUPRO	2260
Db	6721	AGCCAGGGGTGAGGCTGTGGCCAGGCTCATCTTACCGCACCTTGGCGGGGCTACTGCTCT	6780

QY	2261	HiEaenTYrApPProaBpLyBaTSerLeuAryValProLySaBpProIleIeAnThr	2280
Db	6781	CATACTAATGACCTGTGACAGCGAGCTTGAGATGCCCAACCCCGCATCATCAACA	6840
QY	2281	ProValValSerLieserValHisAspAspGluLuleuLeuProArgAlaLeuAspLys	2300
Db	6841	CCCGTGGTACCATCAGGCTCATGAGATGAGAGCTTCGGCCCCGGGCGCTGACAA	6900
QY	2301	ProValThrValGlnPheArgLeuLeuGluLThrGluLugLargThrLysProIleCySerVal	2320
Db	6901	CCCGTACAGGTGCACTTCGGCTCTGTGGAGCAAGAGAGCGGACCAAGCCATCTGTCTC	6960
QY	2321	PheTrpAsnHisSerLileuValSerGlyYThrGlyYTrpSerAlaArgGlyCySerGlu	2340
Db	6961	TTCTTGAAACCATTCMAATCCGTGTCAGTGGCAAGATGTGGTGTGGCGACAGGCTGTGA	7020
QY	2341	ValValAlaPheArgAsnGluSerHisValSerCySerGlnCyAsnHisSmcThrSerPheAla	2360
Db	7021	GTGGTCTTCGGCAATGAGAGCAAGTCAGAGCTGCCAGTGCACCAACATGACAGCTTGCT	7080
QY	2361	ValLeuMetAspValSerArgArgGluAsnGlyValLileuProLeuLysYThrLeuThr	2380
Db	7081	GTGCTCATGAGCAGTTTCTCGCGGGAGAAATGGGAAGATCTCTGCCACTGGAACACTGACA	7140
QY	2381	TyrValAlaLeuGlyValLThrLeuAlaAlaLeuLeuLThrPhePheLeuThrLeu	2400
Db	7141	TACCTGGCTCAGGTGTCACTTGAGTGCCTTCTGTGCACCTTCTTCTCTCATCTCTC	7200
QY	2401	LeuArgGluLeuLeuArgSerAsnGlnHisGlyYLeArgArgAsnLeuThrAlaAlaLeuGly	2420
Db	7201	TTTGGGTATCCCTGGCCCTCAACCAACAGGGATCCGAGTAACTGGACAGCTGCCCTGGGC	7260
QY	2421	LeuAlaGlnLeuValPheLeuLeuGlyYLeAsnGlnAlaAspLeuProPheAlaCySerThr	2440
Db	7261	CTGGCTACGCTGAGTCTTCTCTCTGTGGAAATCAACAGGCTGACTCTCCCTTTTGCTTGACA	7320
QY	2441	ValLleAlaIleLeuLeuHisAspLeuYTrLeuCyThrPheSerTrpAlaLeuLeuGlu	2460
Db	7321	GTCAATTGCCATCTGTGCACCTTCTGTACCTGTGCACCTTTTCTGGGCTCTGCTGGAG	7380
QY	2461	AlaLeuHisIleuLTYrArgAlaAlaLeuThrGluValArgAspValAsnThrGlyProMetArg	2480
Db	7381	GCTTGTGACCTGTGACCGGGCACTCACTGAGGTGGCATGTGCACACCGGCCCATGGC	7440
QY	2481	PheYTrYTrMetLeuGlyYTrpGlyValProAlaPheLleThrGlyLeuAlaValAlaGlyLeu	2500
Db	7441	TTCTACTACATGCTGGGCTGGGGCTGGGCTGCTCTTCACTACAGGGCTAAAGCCGTGGGCTG	7500
QY	2501	AspProGluGlyYTrGlyAsnProAspPheCySerTrpLeuSerLetyrAspThrLeuIle	2520
Db	7501	GACCCCGAGGGCTACGGGAACCTGCACTTGCTGGCTCTCCATCTATGACAGCTCATC	7560
QY	2521	TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuYTrIleu	2540
Db	7561	TGGAGTTTCTGTGGCCCCGGTGGCTTTGGCGTCCGATGAGTGTCTTCCGTGCATCATCCTG	7620
QY	2541	AlaAlaArgAlaSerCySerAlaAlaGlnArgGluYpHeGluLysLysGlyProValSer	2560
Db	7621	GCGGCCCCGGGCTCTGTGTGTGGCCACAGGCGCTTTAGAGAAAGGTCTCTGTCTG	7680
QY	2561	GlyLeuGlnProSerPheAlaValAlaLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu	2580
Db	7681	GGCGCTGAGGCGCTCTTGGCGGTCTCTGTCTGTGTAGGCCCAAGTGGCTGTGGCACTG	7740
QY	2581	LeuSerValAsnSerAspThrLeuLeuPheHisYTrLeuPheAlaThrCyAsnCySrlle	2600
Db	7741	CTCTCTGTCAACAGCGACACCTCTCTCTTCCACTACTCTTTGTGTACTGCAATTGCATC	7800
QY	2601	GlnGlyProPheLlePheLeuSerYTrValValLeuSerYsgLysValaGlySAlaLeu	2620
Db	7801	CAGGAGCCCTTCATCTTCTCTCTTAATGATGATCTTAAGCAAGAGGTCCGAGAAAGCATC	7860
QY	2621	LysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSerThLeu	2640

Db	7861	AAGCTTGCTCAGAGCCGAGGCCAGCCCTGACCTGCTGTGACCAACCAATCCACCTGT	7920
Qy	2641	ThrsersrtyrtaenCyapProserProfyrAlaapaglyAagLeuYrGlInProTyrgly	2660
Db	7921	ACCTGCTCTCAACATGCCCCAGGCCCTACGCAAGATGGCGCGCTGTACACGCCCTACGGA	7980
Qy	2661	AspserAlaglySerLeuHisSerThrsersrArglyLysSerGlnProSerTyrtle	2680
Db	7981	GACTCGCGCGGCTCTCTGTGACAGACCAAGTCGCTCGGCGAAGATCAGCCAGCTACATC	8040
Qy	2681	ProPheLeuLeuArgLglInGlnSerAlaLeuAnProGlynglyProProGlyLeuGly	2700
Db	8041	CCCTCTTGCTGAGAGGAGAGATCCGCACTGAACCTGGCCAGAGGCCCCCTGGCTGGGG	8100
Qy	2701	AspProGlySerLeuPheLeuGlnGlyLysInapGlnGlnHisAspProAspThrAspSer	2720
Db	8101	GATCCAGGACAGCTGTCTCTGGAAAGTCAAGACAGACAGCATGATCTTGACACGGAATCC	8160
Qy	2721	AspserAspLeuSerLeuGlyLysAspGlnSerGlySerTyrtAlaserThrsSerSer	2740
Db	8161	GACAGTGACTGTGCTTGTGAAGAGCAACAGATGGCTCTATGCTTACCACTCACTCA	8220
Qy	2741	AspserGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	2760
Db	8221	GACAGTGAAG	8280
Qy	2761	TrypAspSerLeuLeuGlyProGlyValaGlnArgLeuProLeuHisSerThrProLysAsp	2780
Db	8281	TGGATATGCTGTCTGTGGGCTGTGAGACAGAGAGATGCTGCTGTGACAGTACTCCAGAGAT	8340
Qy	2781	GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGln	2800
Db	8341	GCGGGCCCCAGGGCTCGGCAAGGCCCTCGGCCAGAGACTTTGGGACCAAGCAAAAGAG	8400
Qy	2801	SerSerGlyAnGlyAlaProGlnGlnArgLeuArgGlyLysAnGlyAspAlaLeuSerArg	2820
Db	8401	AGTATGGCCAAAGGGGCCCCGTGAGAGAGCGCTGCGGAGAAATGAGATGCTGTCTTGA	8460
Qy	2821	GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyTleLeuLys	2840
Db	8461	GAGGGGCTCCCTAGAGCCCCCTTCCAGGCTCTTCTGTGCCAGGCTCAAAAGGATCCTTAAG	8520
Qy	2841	LysLysArgLeuProThrTlleSerGlnLysSerSerLeuLeuArgLeuProLeuGlnGln	2860
Db	8521	AAGAAAGTGTGCCCCACATCAAGCAGAGAGACACCTCTCGCGCTCCCCCTGAGGAA	8580
Qy	2861	CysThrGlySerSerArgGlySerSerAlaSerGlnGlySerArgGlyGlyProProPro	2880
Db	8581	TGCACAGGGTCTTCCCGGGGCTCTCTCCGTAGTGAAGGACGCGGGGGGCCCCCTCC	8640
Qy	2881	ArgProProProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	2900
Db	8641	CGCCCAACCGCCCCGAGAGGCTCCAGAGAGCAGCTGAACCGGGGTATATCCCATCGCCATG	8700
Qy	2901	SerLysLysAlaGlyThrValAspGlnAspSerSerGlySerGlnPheLeuPhePheAn	2920
Db	8701	AGCATCAAGGACGAGCAGCGTGAATGAAGATCTGTCAAGGCTCCGAATTTCTCTTTTAAC	8760
Qy	2921	PheLeuHis 2923	
Db	8761	TTCTCTGAT 8769	
RESULT 2			
ID	ABZ42868	standard; DNA, 10531 BP.	
XX	ABZ42868;		
XX	AC		
XX	DT	(first entry)	
XX	04-WAR-2003		
DE	Human GPCR CEUSR2 nucleotide SEQ ID NO:523.		

XX	G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW	G protein-coupled receptor modulator; antibody; immune-related disease;
KW	growth-related disease, cell regeneration-related disease, AIDS; cancer;
KW	immunological-related cell proliferative disease; autoimmune disease;
KW	Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW	osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW	graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW	psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW	mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW	hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW	ulcer; gene; ds.
XX	
XX	Homo sapiens.
OS	
XX	
XX	MO200261087-A2.
PN	
XX	
PD	08-AUG-2002.
XX	
PE	19-DEC-2001; 2001WO-US50107.
XX	
PR	19-DEC-2000; 2000US-257144P.
XX	
PA	(LIFE-) LIFESPAN BIOSCIENCES INC.
XX	
P1	Burmer GC, Roush CL, Brown JP;
XX	
DR	WPI: 2003-046718/04.
XX	
DR	P-PSDB; ABP82018.
XX	
PT	New isolated antigenic peptides e.g., for G protein-coupled receptors
PT	(GPCR), useful for diagnosing and designing drugs for treating
PT	conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
PT	cancer or autoimmune diseases
PS	
PS	Disclosure; Fig 1; 523pp; English.
XX	
CC	The present invention describes antigenic peptides (1) comprising:
CC	(a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC	acids. Also described: (1) an assay for the detection of a particular
CC	G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC	and (2) an isolated antibody having high specificity and high affinity
CC	or avidity for a particular GPCR. (1) can be used as GPCR modulators and
CC	in gene therapy. The antigenic peptides for GPCRs are useful in detecting
CC	an antibody against a particular GPCR, and in the production of specific
CC	antibodies. The peptides and antibodies are also useful for detecting the
CC	presence or absence of corresponding GPCRs. The antigenic peptides for
CC	GPCRs and antibodies are useful for diagnosing and designing drugs for
CC	treating immune-related diseases, growth-related diseases, cell
CC	regeneration-related disease, immunological-related cell proliferative
CC	diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC	atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC	osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC	inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC	disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC	anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC	loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
CC	hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC	any other disorder in which GPCRs are involved. The antibodies may be
CC	used in immunoassays and immunodiagnosis. AB24523 to AB242869 encode
CC	GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC	exemplification of the present invention.
XX	
XX	Sequence 10531 BP; 2038 A; 3353 C; 2598 G; 2142 T; 0 other;
SD	
Alignment Scores:	
Pred. No.:	0
Score:	15545.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	25
	Matches: 10531
	Conservative: 2923
	Mismatches: 0
	Indels: 0
	Gaps: 0

US-09-916-849A-3 (1-2923) x AB242868 (1-10531)
US-09-916-849A-3 (1-2923) x AB242868 (1-10531)



QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db ATCGAGAGCCCGGCGACGGGGTCTCCCTCCCAACGCGCGCGCGCTGCTGCTG 122  
 QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182  
 QY 41 GlySerArgGlyArgGlySerSerGlyValaCysAlaProMetGlyTrpLeuCysProSer 60  
 Db GGCTCCAGGGGAGACAGAGCTCTTGCGGGGGCTTGCGGCCCAATGGGCTGCTGCTGCTG 242  
 QY 61 SerAlaSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80  
 Db TCGAGCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
 QY 81 GlyHisLeuValProHisAlaAspGlyLeuArgValTrpCysProGlySerGlyValaHis 100  
 Db GGCCACTGCTATCCCAACCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362  
 QY 101 IleProLeuProProAlaProGlyGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
 Db ATTCCCTTACCAACAGCTCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 422  
 QY 121 GlyHisLeuSerProGlnGlyValLeuThrLeuProGlyGlyHisProCysLeuVala 140  
 Db GGCCACTTCTCCCAACAGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 482  
 QY 141 ProArgLeuArgCysGlnSerCysGlyLeuValaGlnAlaProGlyLeuArgGlyIle 160  
 Db CCACGCTGAGATGCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542  
 QY 161 ArgSerProGlyGlyLeuSerLeuGlyIleArgArgLeuArgValaAsnThrAlaProGln 180  
 Db AGGTTCACCAAGAGAGTCCCTGGGGGGCTGCGAAAAGAAATTAATACAGCCCCCAG 602  
 QY 181 PheGlnProProSerGlyGlnAlaThrValProGlyLeuGlnProAlaGlyThrProVal 200  
 Db TTTCACGCCCCCAGCTACCAAGGCGACAGTCCGGAACCAAGCCAGCCAGCCCTGCTG 662  
 QY 201 AlaSerLeuArgAlaIleAspProAspGlyGlyValaGlyValArgLeuGlyTrpThrMet 220  
 Db GCATCCCTTACAGGCGCATTCAGCCGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 722  
 QY 221 AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAla 240  
 Db GATGCCCTCTTGAATAGCGGCTCCAAACAGTTCTTCCCTGGAACCAAGTCACTGCTGCA 782  
 QY 241 ValThrThrAlaGlyGlyLeuAspArgGlyLeuThrLysSerThrHisValPheArgValThr 260  
 Db GTAAACCAAGCGGAGGAGGAGTCTGATCGTAGAACCAAGACCAACCTCTTCAAGGGTCAAG 842  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 Db GCGCAGAGACCAAGCGATGCCCGACCAAGAGTGCCTGCTACCTCAACCTTGGTTACT 902  
 QY 281 AspThrAsnAspHisAspProValPheGlyGlnGlnGlyTrpLysGlyLeuSerLeuArgGln 300  
 Db GACACCAATGACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 962  
 QY 301 AsnLeuGlyValaGlyTrpGlyValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db AACCTGAGAGTGGCTATAGAGTCTCACTGTCAGGCGCACCGATGTATGCTCCCTCC 1022  
 QY 321 AsnAlaAsnIleLeuTrpArgLeuLeuGlyGlySerGlyLysSerProSerGlyValaPhe 340  
 Db AATGCCAATATTTCTGACCGCTGCTGAGAGGGGTCTGGGGGGAGCCCTCTGAGAGTCTT 1082  
 QY 341 GlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlyGlyVal 360  
 Db GAGATGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142

QY 361 GluSerTrpGlnLeuThrValGlnAlaSerAspGlnIleArgAspProGlyProArgSer 380  
 Db GAATCTTACAGCTGACGGTAAAGGAGGAGTGCACAGGGTCCGGAGCCGGGCTCTCGAGAT 1202  
 QY 381 ThrThrAlaValaPheLeuSerValGluAspAsnAspAsnAlaProGlnPheSer 400  
 Db ACCACAGCGCTGTTTCTTCTCTGAGAGTGAACAATGAATATATGCCCCCAAGTTTAAAT 1262  
 QY 401 GlnLysArgTrpValaValaGlnValaArgGluAspValThrProGlyValaProValLeuArg 420  
 Db GAGAACGCTATGTGCTCCAGTGAAGAGATGTATCTCCAGGGGCCCAAGTACTCCGA 1322  
 QY 421 ValThrAlaSerAspArgAspLysGlySerAsnAlaValaHisIleTrpSerIleMetSer 440  
 Db GTACACAGCTCGAGATCGAACAAGGGAGCAATGCGTGTGCTCATATAGCATATGAGT 1382  
 QY 441 GlyAsnAlaArgGlyGlyPheTrpLeuAspAlaGlnThrGlyValaLeuAspValaSer 460  
 Db GGCAATGCTGGGGAGACAGTTTATCTGATGAGCCAGACTGAGCTTGATGTGTGAGC 1442  
 QY 461 ProLeuAspTrpGlyThrThrLysGlyTrpThrLeuArgValaArgAlaGlnAspGlyIle 480  
 Db CCTCTTGAATAGAGACAGACCAAGAGATACCTTACGGGTGCGAGCACAGATGTGGC 1502  
 QY 481 ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500  
 Db CGTCCCCCACTCTTAATGTCTGTGGTGTGAGACATGATCTGTGATATCAACGAC 1562  
 QY 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlySerValaProLeu 520  
 Db AATGCCCGCATCTTGTCAGACCCCTTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1622  
 QY 521 GlyTrpLeuValLeuHisValaGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeu 540  
 Db GGCTACCTGCTTCTCAATGCTCAGGCTATCAGCGTATGCTGTGAGCAATCCCGCTG 1682  
 QY 541 GlnTrpArgLeuAlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db GAATACCGCTTGTGCTGGGTGGAGCATGATCCCTTCAACCATCAACATGGCACAGGC 1742  
 QY 561 TrpIleSerValaAlaGlyLeuAspAspArgGlyGlyValaAspPheTrpSerPheGlyVal 580  
 Db TGGATCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1802  
 QY 581 GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db GAACTCCAGACCAAGGAGCTCAGCACTCACTGCTGCGCAAGTCAAGGTGACTGCTG 1862  
 QY 601 LeuAspValaAsnAspAsnAsnProThrPheThrGlnProGlyTrpValaArgLeuAsn 620  
 Db CTGATGTCAACGACCAACATCCAACTTACCCCAACCAAGATACACAGTGGGCTCAAT 1922  
 QY 621 GluAspAlaAlaValaGlyThrSerValaValaThrValSerAlaValaAspArgAspAlaHis 640  
 Db GAGATGAGCTGTGGGACCAAGGCTGTGAGCGGTGAGCTGTGAGCGTGAATCTCAT 1982  
 QY 641 SerValaIleThrTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db AGTGCATCACTTCAACATCAACAGTGGCAATCTGCAAAACGCTTCTCATCAACCAAGC 2042  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTrpLysLeuGlyLeuArgGln 680  
 Db CAATGTGTGTGGGTGTGATCTTCCCTGCGCACTGAGCTCAAACTTGAAGGGGAG 2102  
 QY 681 TyrValLeuAlaValaThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValaVal 700  
 Db TATGTTGGCTTTACCGCTCCGATGACCTGGCGAGACCGGCAAGATGTGTG 2162  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisIleTrpValaAsn 720  
 Db AATGTCACGACGCAACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2222  
 QY 721 ValaGlnLeuAspArgProAlaGlyThrThrValaValLeuIleSerAlaThrAspGlyLeu 740

D	2223	GTATATGAGCACCGCCGCGGACGACCAACGCTGGCTGATCAGCGCCACGAGAGAGAC	2282
Q	741	ThrglgluAanlaArglleThrTyrPheMetgluAseSerlleProglInpHatrglle	760
D	2283	ACAGGTGAAATGCCCGCATCACTTCAATGAGAGACAAATCCCACTTCGCGATC	2342
Q	761	AspAlaAspThrArglalaValThrThrglnalagluleuAspTyrGluluaSpGlInvalSer	780
D	2243	GATCGACGACGCGGGGCTGTCAACCAACGAGCTGAGCTACGAAAGCAAGTGTCT	2402
Q	781	TyrThrleuAlaIleThrAlaArgAspanglylleProglInlySseAspThrThrTyr	800
D	2403	TACACCCCTGGCCATTACGTCTCGGAGCAATGSCATTCGCCAAGAGTCGACCACTTAC	2462
Q	801	LeugluileuValaAspAspValaAspAspAlaProglInpHeuArgAspSerTyr	820
D	2463	CTGAGATCTCTGTGAGACGAGTGAATGACATGCCCTCAGTTCCTGGAGACTCCCTAC	2522
Q	821	GlnGlYserValTyrGluluaSpValProProPheThrSerValleuglnIleSerAlaThr	840
D	2523	CAGGCAAGTGTCTATGAGAGATGTGCACCTTCACTAGCGTCTGCAAGTCTCAGCCACT	2582
Q	841	AspArgAspSerArglyleuAanglyArgValPheTyrThrPheglnglylAspAspGly	860
D	2583	GATGCTGATTCTGACTTAATGCGAGGCTCTTCAACCTTCCAAAGAGGCGACGATGGA	2642
Q	861	AspAlYAspPheIleValgluSerThrserylleValArgThrleuArgArgleuAsp	880
D	2643	GACCGTACTTATTTGTTGATCCACGTCAGGCACTCGGCAAGCCTACGAGGCTGGAT	2702
Q	881	ArggluAanValaaglInTyrValleuArglaIaTyrAlaValaAspLyglYmetProPro	900
D	2703	CGAAGAAACGTGGCCAGTATGTCTGGGCGCATATGCAATGAGCAAGGGGATCCCCCA	2762
Q	901	AlaArgThrProMetgluValThrValThrValleuAspValaAspAspProProVal	920
D	2763	GCCCGCACCTATGAGAGTACAGTCACTGTGTGATGATGAAATGCAATCCCGCTGC	2822
Q	921	PhegluInaSpGluluaSpValaPheValgluGluluaSerProIleGlYleuAlaVal	940
D	2823	TTTGAAGAGATGATGATTGATGTGTGTGTGAAAGAGAACGCCCATTTGGGCTAGCCGTG	2882
Q	941	AlaArgValThrAlaThrAspProAspGlulglYThrAanlaaglInIemetTyrGlInle	960
D	2883	GCCCGGTGCACAGCCATGACCCCGATGAAAGCAACATGCCAGATTATCACAATT	2942
Q	961	ValgluGlYAsnIleProglInValPheglInleuAspIlePheSerGlYgluLeuThrAla	980
D	2943	GTGAGGGCAACATCCCTGAGGCTCTCCAGCTGACATCTTCTCGGGGAGCTACAGCC	3002
Q	981	LeuValaAspLeuAspTyrGluluaSpArgProglInTyrValleuValIleGlInlaIeThSer	1000
D	3003	CTGGTAAAGCTTAAGCTACGAGGACCGGCTCTAGTACGTCTGTGTATTCAGGCCACGTCA	3062
Q	1001	AlaProleuValaSerArgAlaThrValIleValArgleuLeuAspArgAspAspPro	1020
D	3063	GCTCTCTGTGTAGCGCGGCTACAGTCCAGTCCGCTCTTGGACCGGAATGAAACAACCA	3122
Q	1021	ProValleuGlYAsnPhelgluIleleuPheAspAsnTyrValThrAanArgSseSer	1040
D	3123	CGAGTGGCGCAATCTTGAATCTTTTCAACAATATGTCACCAATCGCTCAAGCAAGC	3182
Q	1041	PheProglYglYAlaIleGlYArgValProAlaHisAspProAspIleSseAspSerleu	1060
D	3183	TTCCCTGGGGGTGCATTGGCCGAGTACCTGCCATGACCTGATATCTCAGATAGCTCG	3242
Q	1061	ThrTyrSerPhegluArgglYAsnGlYleuSerleuValleuAanlaIleSerThrGlY	1080
D	3243	ACTTACAGCTTGTGACGGGAAATGAATCAAGCTGTCTGTCCAAATGCTTCCACGGGT	3302
Q	1081	GluleuYleuSerArgAlaIleuAspAanArgProleuGlulAlaIleMetSerVal	1100
D			
D	3303	GAGCTGAGCTTAAGCCGCGCATGACAAACAACGCGCTCTGGAGGCAATCATGACGTG	3362
Q	1101	LeuValaSerAspGlYValHisSerValThrAlaGlYCyAlaIleuArgValThrIleIle	1120
D	3363	CTGTGTCAACAGCGGATACAGCTGACCGCCCAAGTGGCGCTGGGTGATCAATCATC	3422
Q	1121	ThrAspGluMetLeuThrHisSerIleThrleuArgleuGluluaPheSerProgluArg	1140
D	3423	ACCGATGAGATGCTCAACCAAGCATCAAGCTGCGCTGGAGGACATGTCAACCGAGCGC	3482
Q	1141	PheleuSerProleuLeuGlYleuPheIleGlInlaValAlaIleThrleuAlaIleThrPro	1160
D	3483	TTCTCTGACCACTGTCTAAGCCCTTTCATCCAGGCGGTGGCCGACGCTGGCGACGCCA	3542
Q	1161	ProAspHisValValaValaPheAsnValaIaArgAspThrAspAlaProglYglYHISile	1180
D	3543	CCGAGCACGTGTGTGCTTCAACGTACAGCGGAGACAGAACCCCGGGGGCCACATC	3602
Q	1181	LeuAanValaSerleuSerValglYglInProProglYProglYglYProProPheleu	1200
D	3603	CTCAAGTGAAGCTGTGGGTGGGCGAGCCGCGGAGCGCCGCGGGGCGGCCCTTCTCG	3662
Q	1201	ProSerGluluaPheglnglylueuTyrleuAanArgSerleuLeuThrAlaIleSer	1220
D	3663	CCCTTGAAGGACCTTGACGAGAGCGCTTATACCTCAACCGAGCCTGTGACGGCCATCTCG	3722
Q	1221	AlaGlInArgValleuProPheAspAspAlaIleCyAlaArgglulProCygluAanTyr	1240
D	3723	GCAACGCGGTGCTGCCCTTCAACGACAACTGTGCTGGAGGCCCTGGAGAACTAC	3782
Q	1241	MetArgCyAlaSerleuValleuArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260
D	3783	ATGCGCTGGCGTGGTGTGCTGCCTCGACTCCTCGCGCCCTTCAATCCGCTCTCTCC	3842
Q	1261	ValleuPheArgProIleHisProValaIglYglYleuArgCyAlaArgCyAspProProglYpHe	1280
D	3843	GTGCTCTTCGCGCCATCAACCCCGTGGAGGGCTGGCGCTGGCCCGCCGCGCTTC	3902
Q	1281	ThrGlYAspTyrCyaglInThrGluluaIleAspLeuCyTyrSerArgProCyglYProHis	1300
D	3903	ACGGGTGATCTACCTGACAGACCAAGATGACCTGTGTAATCGCGGCGCTGTGGCCCCCAG	3962
Q	1301	GlYArgCyAlaArgSerArgglYglYlYThrCyAlaArgAspGlYlYThrGlY	1320
D	3963	GGCGCTGGCGGCGCGGAGGGCGGCTACACCTGCTGTGTGTATGCTACAGGGGT	4022
Q	1321	GlulHisCygluValaSerAlaArgSerArgYArgCyThrProglYValCyAlaYAsnAngly	1340
D	4023	GAGCACTGTAGGTGAGTGTGCTGAGGCGCTTGACCCCGGGGTGTGCAAGAAATGGG	4082
Q	1341	GlYThrCyAlaAsnleuLeuValglYglYpHeYCyAspCyAspSerArgYAspPhe	1360
D	4083	GGCACTGTGTCAACTGCTGTGTGGCGGTTTCAATGTGGAATGGCCATCTGGAGACTTC	4142
Q	1361	GlulYAspProTyrCygluValThrThrArgSerPheProAlaHisSerPheIleThrPhe	1380
D	4143	GAGAAACCTTACCTGACGAGGTGACCAAGCGCAGCTTCCCGGCCCATCTTCAATCACTTT	4202
Q	1381	ArgGlYleuArggluArgPheHisPheThrleuAlaIleuSerPheAlaThrlygluArg	1400
D	4203	CGCGGCTGGCGAGGCTTTCATCTTCACTTGAGCCCTTCTGTTTCCCAAGAGAGCGC	4262
Q	1401	AspGlYleuLeuLeuTyrAanglyArgPheAsnGlYHisAspPheValaIleuGlul	1420
D	4263	GACGGTGTGTGTGATGAGAGGCTTCAATGAAGAACATGACTTGTGGCCCTCGAG	4322
Q	1421	ValIleGlIngluValaIleleuThrPheSerAlaIglYgluSerThrThrValaIleSer	1440
D	4323	GTATTCAGAGAGAGGTCCAGCTCACTTCTGCGAGGGAAGTCAACACAGGTGTCC	4382
Q	1441	ProPheValaProglYglYValaSerAspGlYglInThrPheThrValglInleuYsTyrTyr	1460
D	4383	CAATGTGTGCGGAGAGATCATGTATGCGCAATGTCATACGCTGCAAGTGAATCTAC	4442

QY	1461	AsnLysPLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnLysValAla	1480
DB	4443	AAAPAGGACATGTTGGGTGACAGACGAGGCTCCACAGGGCCCATCAGAGCAGAGGTGGCT	4502
QY	1481	ValValThrValAlaAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly	1500
DB	4503	GTGGTAACTGGAGTGGCTGTGACACAGAGTGGCTTGGCTTGGATCTGTCTGGAGC	4562
QY	1501	AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysSerLeuAspLeuThr	1520
DB	4563	AACTACTCCTGTGTGGTCCAGGGGACCCAGGTGGGACAGAAAGTCTTGATCTGACG	4622
QY	1521	GlyProLeuLeuLeuGlyGlyValProAspLeuProGlnSerPheProValArgMetArg	1540
DB	4623	GGGCCCCCTGCTACAGCGGGGTGGCTGACCTGGCCCAAGAGCTTCCAGCTCCAAATGGG	4682
QY	1541	GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp	1560
DB	4683	CAGTTGTGGGCTGACATGCGGAACCTGACAGTGAAGCCGGCACATGACATGGCTGAC	4742
QY	1561	PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysIleValAsnValCysAspSer	1580
DB	4743	TTCATTGGCCAAATGGCACCGTGGCTGGCTGGCTGGCCAAAGAAAGTGTGTGACAGC	4802
QY	1581	AsnThrCysHisIleAsnGlyGlyThrCysValAsnGlnTrpAspAlaPheSerCysGluCys	1600
DB	4803	AAACACTGCGCAATGGGGGCACTTGGTGAACCAAGTGGGACCGGTTCAGCTCGAGTGC	4862
QY	1601	ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu	1620
DB	4863	CCCCCTGGGCTTTGGGGGACAGACCTGGCCAGGAAATGGCCAAATCCACAGCACCTTCCG	4922
QY	1621	GlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyrLeu	1640
DB	4923	GGCAGCAGCTGGTGGCTGGCATGGCTCTGGCTGCCATCTCCAACTCGTGTACTC	4982
QY	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660
DB	4983	AGCTCATGTCTCGACGCGCCAGGCGGACGAGTGTCTGTGACGCGCATCACAGAGGGG	5042
QY	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlyGlyThr	1680
DB	5043	CGAGACACATCACCTTACAGCTACAGAGAGGCCCATGTATGCTGACGTTGAGGGGACAC	5102
QY	1681	GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTrpHis	1700
DB	5103	GGGCTTCAGGGCTCTCTCTCGCTGGAGCCAGGCCGGGCCCAATGACGGTATCTGGCAC	5162
QY	1701	HisAlaGlnLeuAlaLeuGlyValAspGlyGlyProGlyHisAlaIleLeuSerPheAsp	1720
DB	5163	CATGCACACTGGCATGGAGGACAGCGGGGGGCTGGCCATGCCATTCGTCTTCGAT	5222
QY	1721	TyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGlyLeuHisIleLeuSer	1740
DB	5223	TATGGGACAGAGAGACAGAGGACACTGGGCCCCCGGCTGTGATGTCTGACCTGAGC	5282
QY	1741	AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaAspGlyPheArgGly	1760
DB	5283	AAACATAACAGTGGGCGGAATACCTGGGCCAGCGGGGGTGGCCGTGGCTTGGGGG	5342
QY	1761	CysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsnSerLeuAspProSer	1780
DB	5343	TGTTTGCAGAGGTGTGGGTGAGCAGATACGCCAGAGGGGGTTAAACGCTTGATCCAC	5402
QY	1781	HisGlyGlnSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
DB	5403	CATGGGAGAGCATCACTGGAGCAAGGCTGTAGCTGTGACCTTGTGTATCTAACC	5462
QY	1801	ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp	1820
DB	5463	CCGTGTCTGTAAAGCTATTTGACAAAGACTGGGACAGCTAATCTCTGACAGCTGTGAT	5522
<hr/>			
QY	1821	ProGlyTyrTyrGlyValAspAsnCysThrAsnValCysAspLeuAsnProCysGlyHisGln	1840
DB	5523	CCAGGTACTAATGTGACAACTGTACTATATGTGTGACTTGAACCCGTGTGAGCACAG	5582
QY	1841	SerValCysThrArgLysPheSerAlaProHisGlyTyrThrCysGlyCysProProAsn	1860
DB	5583	TCTGTGTAACTCCGCAAGGCCAGGTGGCCCCCATGTGTATATCTGCAAGTGTCCCCAAAT	5642
QY	1861	TyrLeuGlyProTyrCysGlyLysThrArgIleAspGlnProCysProArgGlyTrpTrpGly	1880
DB	5643	TACCTTGGGCACTTACTGTAGACAGAGATTGACAGACCTTGTCTCCGTGGGTGGAG	5702
QY	1881	HisProThrCysGlyProCysAsnCysAspValSerLysGlyPheAspProAspCysAsn	1900
DB	5703	CATCCCAATGTGGGCCCATGTGACATGTGATGTCAAGAAAGGCTTTGACCCAGCTGCAAC	5762
QY	1901	LysThrSerGlyGlyCysHisIleCysLysGlyAsnHisIleTyrArgProProGlySerProThr	1920
DB	5763	AAAGCAAGCGGGAAGTGCATGTGCAAGAGAAACACTACCGGCCCCCAGGAGCCCAAC	5822
QY	1921	CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGluAsp	1940
DB	5823	TGCCTTGTGTGATGATGCTTACCCCAAGGCTCTTGTCCAGAGTCTGTGACCTTAGAT	5882
QY	1941	GlyGlnCysProCysLysAspProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro	1960
DB	5883	GGCCAGTGTCCATGCAAGCAGAGTGTATCGGGCGCTAGTGTGACCGCTGTGACAACTT	5942
QY	1961	PheAlaGlnValThrThrAsnGlyCysGlyLysValAsnTyrAspSerCysProArgAlaIle	1980
DB	5943	TTTGGTAGGTATCAACCAATGGCTGTGAAGTAATTAAGACGTGCCCAAGACGACGANT	6002
QY	1981	GlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaAlaAlaProCysPro	2000
DB	6003	GAGCTGGGATGTGGTGGCCCGGATACCGGCTTGGGGCTGCTGTCTGTCTCTCTCT	6062
QY	2001	LysGlySerPheGlyThrAlaValArgHisCysAspGlyHisArgGlyTyrLeuProPro	2020
DB	6063	AAAGCTCTTTGGGACATGTGTGGCTGTGCGCATGTGTATGACACAGGGGGTGGCTCC	6122
QY	2021	AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLysGlyPheAlaGlyLysLeu	2040
DB	6123	AACTCTTCACTGACATGATCCATCACCTTCTCAAGATGAAGGGCTTGTGTAGGGCTA	6182
QY	2041	GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArg	2060
DB	6183	CAGCGGAATGATGACAGGCTTGAATCAGGGGCTCTCCAGAGCTAGCCCTGTCTTGGCC	6242
QY	2061	AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu	2080
DB	6243	AACTCCACGACACACAGCTGGCTTACTTCCGACAGCGATCTCAAGGTGGCTTACAGCTG	6302
QY	2081	AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaPheGln	2100
DB	6303	GCCACGCGGCTGTGGCCCAAGAGACCCAGCGGGCTTGTGGCTGTCTGCAACAG	6362
QY	2101	AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn	2120
DB	6363	GACGTGCACTTCACTGAAATCTGTGGGTGGGCAACGCGCTCTCTGACAGACCCAC	6422
QY	2121	LysArgHisIleTrpGlyLeuLeuIleGlnGlnThrGlnGlyGlyThrAlaTrpLeuLeuGlnHis	2140
DB	6423	AAAGCGCACTGGAGCTGATATCAGAGACAGAGGGTGGCACCGCTGTCTGTCCAGAC	6482
QY	2141	TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisIleTyrLeuSerProPhe	2160
DB	6483	TATGAGGCTTACGCAAGTGCCTTGGCCCAAGACATGCGGCACTTACTTAAGCCCTTC	6542
QY	2161	ThrIleValThrProAsnIleValIleSerValArgLeuAspLysGlyAsnPheAla	2180
DB	6543	ACACATGTCAGGCCCAACATTTGATCTCGTATGGGCTTGGACAAAGGAACTTGTCT	6602
QY	2181	GlyAlaLysLeuProArgTyrGlyAlaLeuArgGlyGlnGlnProProAspLeuGlnThr	2200

Db 6603 GGGGCCAAGCTGCCCCCTCAGAGCCCTGCTGGGAGAGCCGCCGAGCTTGAAGCA 6662  
Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProIleGly 2220  
Db 6663 ACGATCTTCTGCTGAGTCTGTCTTTCAGAGAGAGCCGCCCTGGTCAAGGCCGAGGC 6722  
Qy 2221 ProGlyValIleGluInuProGluInuLeuValArgArgGluArgArgIleProGluLeu 2240  
Db 6723 CCGGAGAGAGCCCGAGGAGAGAGAGCTGGCAAGGCGGCAAGCGAGCCGAGGCTG 6782  
Qy 2241 SerGlnGlyGluValValIleSerValIleIleThrArgThrIleuValArgIleuPro 2260  
Db 6783 AGCCAGGGTAGGCTGTGGCCAGGCTCATCTACCGCACCCCTGGCCGAGCTACGCTC 6842  
Qy 2261 HisAsnTrpAspProAspIleArgSerLeuArgValProIleIleAsnThr 2280  
Db 6843 CATACCTATACCTGTACAGAGCCGAGCTTGAAGTCCCAAGGCCGATCATACAGCA 6902  
Qy 2281 ProValValSerIleSerValIleAspAspGluGluLeuLeuProArgAlaLeuAspIle 2300  
Db 6903 CCGGTGGTAGACATCAAGCTTCATGATGATGAGAGCTTCTGCCCCGGGCCCTGAGCAA 6962  
Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluValArgThrIleProIleCysVal 2320  
Db 6963 CCGGTACGGTGCAGTCCCGCTGCTGAGACAGAGAGCGGACCAAGCCATCTGTCTC 7022  
Qy 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyValIleTrpSerAlaArgGlyCysGlu 2340  
Db 7023 TTCTGGAGAACATTCATCTCTGCTGCTGAGCAAGAGTGGCTGTGCTGGCCAGAGCTGAA 7082  
Qy 2341 ValValPheArgArgGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
Db 7083 GTGCTCTCCGCAATGAGAGCCAGCTCAGCTGCCAGGCAACCATGAGAGCTTGGCT 7142  
Qy 2361 ValLeuMetAspValIleSerArgArgGluAsnGlyGluIleuProLeuArgThrLeuThr 2380  
Db 7143 GTGCTACAGAGAGCTTCTCGCGGAGAGATGGGAGATCTGCTGCACTGAAGACATGACA 7202  
Qy 2381 TyrValAlaLeuGlyValIleThrLeuAlaAlaLeuLeuThrPhePheLeuThrLeu 2400  
Db 7203 TAGGCTGCTGAGGTGTCACCTTGGCTGCCCTTCTGCTCACTTCTTCTTCTCACTCTC 7262  
Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgLeuLeuThrAlaAlaLeuGly 2420  
Db 7263 TTGGGTATCTCGCTCAACCAACAGGCACTCCGAGTAACCTGACAGCTGCGCTGGGC 7222  
Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaPheLeuProPheAlaCysThr 2440  
Db 7223 CTGGCTCAGGCTGGCTTCTCTCTGAGATCAACAGGCTGACCTCTTGTGCTGACA 7382  
Qy 2441 ValIleAlaIleLeuLeuHisPheLeuThrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460  
Db 7383 GTGATTCGCATTCGCTGCACTTCTGATCCTCTGCACTTTCCTGGCTCTGCTGGAG 7442  
Qy 2461 AlaLeuHisLeuThrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
Db 7443 GCGCTTGACCTGACCGGCACTCACTGAGGTGCGCATGCAACACCGGCCCATGCGC 7502  
Qy 2481 PheTyrTrpMetLeuGlyTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
Db 7503 TTCTACTACATGCTGGGCTGGGCGGTGGCTGCTTCACTACAGGGCTAGCGGAGGCTG 7562  
Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520  
Db 7563 GACCCCGAGGGCTACGGGAACTCTGACTTGTGCTGCTTCATCTATGACAGCTCATC 7622  
Qy 2521 ThrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrpIleLeu 2540  
Db 7623 TGGAGTTTGTGGCCCGGTGGCTTGTGCTGCTGATGATGTCTTCCGTGATCATCTG 7682  
Qy 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGluValGlyProValSer 2560

Db 7683 GCGGCCCGGGCTCTCTGTGCTGCCAGCGGCAAGGCTTGAAGAAAGTCTGTCTCG 7742  
Qy 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
Db 7743 GGCCTTGACGCTCTTGTGGCGTCTTCTGTGCTGAGCGCACCTGTGCTGTGGCATG 7802  
Qy 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
Db 7803 CTCTGTGTCACAGAGACACCTCTCTTCCACTACTCTTGTGCTACTGCAATTGCATC 7862  
Qy 2601 GlnGlyProPheIlePheLeuSerTyrValIleLeuSerGlyValArgValAlaLeu 2620  
Db 7863 GAGGCGCCCTTCACTTCTCTCTATGTGTGTGTACAGAGAGTCCGGAACACTTC 7922  
Qy 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
Db 7923 AAGCTTGCTGAGCGCCAGGCCAGCCCTGACCTGTCTGACCAAGATCCACCTG 7982  
Qy 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
Db 7983 ACCTGCTCTACACTGCCCCAGCCCTACGAGATGGGCGGCTGTACAGCCCTACGGA 8042  
Qy 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
Db 8043 GACTGGCCGGCTCTCTGACAGCACAGTGGCTGGGCGAGAGTCAAGCCAGCTACATC 8102  
Qy 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
Db 8103 CCTCTTGTGAGGAGAGAGTCCGCACTGAACTTGGCCAAAGGCCCTCGCTGGGG 8162  
Qy 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnHisAspProAspThrAspSer 2720  
Db 8163 GATCCAGGAGGCTGTCTTCTGAGAGTCAAGACAGCAAGCATGATCTGACACAGCACTCC 8222  
Qy 2721 AspSerAspLeuSerLeuGluGluAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
Db 8223 GACAGTACCTGTCTTGAAGACAGCAAGATGGCTCTTATGCTCTTACCACTCATCA 8282  
Qy 2741 AspSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760  
Db 8283 GACAGTACAG 8342  
Qy 2761 TrpAspSerLeuLeuGlyProGlyValArgLeuProLeuHisSerThrProLysAsp 2780  
Db 8343 TGGATAGCTGCTGTGGGCTGGAGACAGAGACATGCCCCCTGCAAGTCTCCAAAGAT 8402  
Qy 2781 GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800  
Db 8403 GGGGGCCAGGGCTGTGGCAAGGCCCTTGGCCAGAGACTTTGGGACCAAGCAAAAGG 8462  
Qy 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
Db 8463 AGTATGGGCAAGGGGCCCTTGAAGAGGGCTGCGGAGAAATGGAGATGCCCTGTCTGA 8522  
Qy 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyIleLeuLys 2840  
Db 8523 GAGGGTTCCTTGGGCCCTTCAAGGCTTCTTGTGCCAGCCCTCAAAAGGCAATCTTAA 8582  
Qy 2841 LysLysCysLeuProThrIleSerGluLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
Db 8583 AAGAGTGTCTGCCACCACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642  
Qy 2861 CysThrGlySerSerArgLysSerSerAlaSerGluGlySerArgGlyValProProPro 2880  
Db 8643 TGCACAGGGTCTTCCCGGGGCTCTCCGCTATGAGGAGCGCGGGCGGCCCTTCCC 8702  
Qy 2881 ArgProProProArgGluSerLeuGlnGluGluGluGluGluGluGluGluGluGlu 2900  
Db 8703 CGCCCAAGCCCGGAGAGGCTCCAGAGAGAGCTGAGAGAGGAGTCAAGCCATCGCAAG 8762  
Qy 2901 SerIleLysAlaGlyThrValAspGluAspSerSerGlySerGluPheLeuPheAsn 2920  
Db 8763 AGCATACAGGAGGAGGAGGAGTGTAGAGACTGTAGAGGCTCCGAATTTCTCTTTAAC 8822

QY 2921 PheLeuHis 2923  
 Db 8823 TTCCTGCAT 8831

## RESULT 3

AS11677  
 ID AS11677 standard, cDNA; 8871 BP.

XX AS11677;

XX 24-OCT-2001 (first entry)

XX Human Flamingo cDNA sequence.

XX Flamingo; human; splice variant; G-protein coupled receptor; diabetes;  
 signal transduction pathway; bacterial; fungal; viral; protozoan; cancer;  
 anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; HIV;  
 obesity; hypotension; hypertension; urinary retention; angina pectoris;  
 myocardial infarction; stroke; ulcer; allergy; anxiety; schizophrenia;  
 benign prostatic hyperplasia; psychotic disorder; neurological disorder;  
 manic depression; delirium; dementia; severe mental retardation; ss;  
 Huntington's disease; Gilles de la Tourette's syndrome; antibacterial;  
 antifungal; antiviral; antiparasitic; anti-HIV; anorectic; antianorectic;  
 antiParkinsonian; cariant; cerebroprotective; neuroprotective;  
 antidepressant; anticonvulsant; antisense therapy; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..8771

XX /tag= a /product= "Human Flamingo protein #1"

XX WO200161003-A1.

XX 23-ANG-2001.

XX 19-FEB-2001; 2001WO-GB00680.

XX 19-FEB-2000; 2000GB-0004196.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Testa TT;

XX WPI; 2001-502792/55.

XX P-SDB; AAU07053.

XX An isolated Flamingo polypeptide useful for treating diseases such as  
 HIV, cancer, asthma, Parkinson's disease, acute heart failure,  
 osteoporosis -

XX Claim 5; Page 25-28; 66pp; English.

XX The sequence represents a cDNA which encodes a human Flamingo  
 polypeptide. Flamingo is a member of the G-protein coupled receptor  
 family, which is involved in signal transduction pathways. By screening  
 to identify compounds that stimulate or inhibit the function or level of  
 the protein, treatments can be developed for various diseases and  
 bacterial, fungal, protozoan and viral infections, including HIV, cancer,  
 diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease, acute  
 heart failure, hypotension, hypertension, urinary retention, angina  
 pectoris, myocardial infarction, stroke, ulcers, allergies and benign  
 prostatic hyperplasia. Also treatable are psychotic and neurological  
 disorders such as anxiety, schizophrenia, manic depression, delirium,  
 dementia, severe mental retardation, Huntington's disease and Gilles de  
 la Tourette's syndrome.

XX Sequence 8871 BP; 1712 A; 2859 C; 2587 G; 1713 T; 0 other;

XX Alignment Scores:

XX Pred. No.: 0 Length: 8871

Score: 15518.50 Matches: 2923  
 Percent Similarity: 98.88  
 Best Local Similarity: 98.88  
 Query Match: 99.83  
 DB: 22 Indels: 33  
 Gaps: 1

US-09-916-849a-3 (1-2923) x AS11677 (1-8871)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 Db 1 ATCCGAGCCCGCCGACCGGCTCCCTCCCAAGCCGCGCGCTGCTGCTG 60  
 QY 21 LeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 Db 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 QY 41 GlySerArgGlyArgGlySerSerGlyValAlaProMetGlyTyrLeuCysProSer 60  
 Db 121 GGATCCAGGAGGAGGAGGCTTCCGAGGCTGCTGCTGCTGCTGCTGCTGCT 180  
 QY 61 SerAlaSerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyTyrGluLeuThr 80  
 Db 181 TGAAGCTCAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 81 GlyHisLeuValProHisHisAspGlyLeuArgValTyrCysProGluSerGluAlaHis 100  
 Db 241 GGCCACCTGGTACCCCAACGATGCGTGAAGGTTGGTTCAGAAATCCGAGGCCCAT 300  
 QY 101 IleProLeuProProAlaProGluGlyCysProTyrSerCysArgLeuLeuGlyIleGly 120  
 Db 301 ATTCCTCCACACAGCTCTGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 QY 121 GlyHisLeuSerProGluGlyLeuSerThrLeuProGluGluHisProCysLeuValAla 140  
 Db 361 GGCCACCTTTTCCCAAGGAGGAGCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 QY 141 ProArgLeuArgCysGlnSerCysAlaSerLeuAlaGlnAlaProGlyLeuArgAlaGly 160  
 Db 421 CCACGGCTCAGATGCAAGTCTGCAAGTGGCAAGGCTGCTGCTGCTGCTGCTGCT 480  
 QY 161 ArgSerProGluGluSerLeuGlyValArgArgValAsnThrAlaProGln 180  
 Db 481 AGGTCAACGAGAGATCCCTGGTGGGCTGCAAGAAAGAAATGTAAATACACCCCC 540  
 QY 181 PheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyTyrProVal 200  
 Db 541 TTCCAGCCCGCCAGCTACCAAGCCCAAGTCCGAGAAACAGCCAGGACCCCTGTT 600  
 QY 201 AlaSerLeuArgAlaIleAspProAspGluGlyAlaGlyArgLeuGluTyrThrMet 220  
 Db 601 GCATCCCTGAGGGCCATGCAACCCGAGCAGAGGTGAGGAGTCACTGAGATCACCATG 660  
 QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAla 240  
 Db 661 GATGCCCTTTGATAGCCGCTCCACCAAGTCTTCTCCGAGCCCACTGCTGCTGCT 720  
 QY 241 ValThrThrAlaGluGluLeuAspArgGluThrThrLeuSerThrHisValPheArgValThr 260  
 Db 721 GTAACCAACGCGAGGAGCTGATGTGAGACCAAGACACCACTCTTCAAGGCTCAG 780  
 QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrTyrLeuValThr 280  
 Db 781 GGGCAGGACCAACGGATCCCGGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCT 840  
 QY 281 AspThrAsnAspHisAspProValPheGluGlnGlnGluTyrGlyGluSerLeuArgGlu 300  
 Db 841 GACACCAATGACCAATGACCTCTGTTCCAGACGAGGATACAGAGAGCTCAGGAG 900  
 QY 301 AsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProPro 320  
 Db 901 AACCTGAGAGTGGCTATGAGGTGCTCACTGCAAGGACGAGATGATGATCCCTCC 960  
 QY 321 AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlySerProSerGluValPhe 340

```

Db 961 AATGCCAATATTCGTACCGCTCTGAGGGGCTGAGGGCAACCCCTCTGAAAGTCTTT 1020
Qy 341 GtUileaspProargSerGlyValIleargThrArgGlyProValAaPaArgGluVal 360
Db 1021 GAGATCAACCTCGCTCTGGGGTATTCGAACCCGTGGCCCTGTGATCGGAAAGAGTG 1080
Qy 361 GluSerTyrglnLeuThrValGluAleserApGlnIyArGaapProGlyProArgSer 380
Db 1081 GAATCCACCAAGTGAAGGTGAGCAAGTGAACAGAGGTGGAGCCGGGCTCTCGAGT 1140
Qy 381 ThrThrIalaIaValPheLeuSerValGluAspAspAspAspAaPaProGlnPheSer 400
Db 1141 ACCACAGCCGCTGTTCCTTCTGTGAGAGATGACATGATATATGCCCCCACTTTAGT 1200
Qy 401 GluUyArGlyValValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
Db 1201 GAGAGAGGCTATATGTGTCTGAGTGAAGGAGATGTGACTCGAGGGGCCCAAGTCTCGA 1260
Qy 421 ValThrAlaserAaPaArgApGlySerAaAaIaValAlHsTySerIleMetSer 440
Db 1261 GTCAACAGCTCGATTCAGACAGCAAGGGAGCAATGCCGTGTGCACATAGCATGAGT 1320
Qy 441 GlyAspAaIaArgGlyGlnPheTyrlLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
Db 1321 GGCATATCTCGGGAGAGTTTATCTGATGCCAGCTGAGCTGAGTGTGAGTGC 1380
Qy 461 ProLeuAspTyrglnThrThrTyrlGlyGlyTyrlLeuArgValAaGlaGlnAspGlyGly 480
Db 1381 CCTCTTACATATGAGACAGCAAGAGATGACCTTACGGGTGCAAGACAGATGCTGCGC 1440
Qy 481 ArgProProLeuSerAaAaValSerGlyLeuValThrValGlnValLeuAspIleAspAsp 500
Db 1441 CGTCCCCCACTCTAATGTCTGTGGCTGTGTGACAGTACAGTCTCGATATCAAGAC 1500
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlySerValProLeu 520
Db 1501 AATCCCCCATCTTGTGACAGACCCCTTTCAGAGCTATGTCTCGAGAGAGGTCCCTTA 1560
Qy 521 GlyTyrlLeuValLeuHsValGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgLeu 540
Db 1561 GGTACACCGGTTTCCATATGTCAGAGCTATGCAAGCTGATGTGTGCAATGCCGCTG 1620
Qy 541 GluTyrlArgLeuAlaGlyValGlyHsAspPheProPheThrIleAsnAspGlyTyrlGly 560
Db 1621 GAATACCGCCTTGTGTGGGTGAGCATGACTTCCCTTCAACATCAACATGGGACAGGC 1680
Qy 561 TrpIleSerValIalaIaGluLeuAspArgGluGluValAspPheTyrlSerPheGlyVal 580
Db 1681 TGAATCTCTGTGGCTGTGACTGAACTGAGCCGGAGAGATGATTTCTTCAAGCTTTGGGTA 1740
Qy 581 GluAlaArgAspHsGlyTyrlProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1741 GAAGCTCGAGACATAGGCACTCCAGCACTCACTCGCTCGGCAAGTCAAGGTGACTGTC 1800
Qy 601 LeuAspValAlaAsnAspAsnAspProThrPheThrGlnProGlyTyrlThrValArgLeuAsn 620
Db 1801 CTGATATCAACGACCAACATCAACCTTACCAACCAAGTACACAGTCCGCTCAAT 1860
Qy 621 GluAspAlaIaValGlyTyrlSerValValThrValSerAlaValAlaAspArgAspAlaHs 640
Db 1861 GAGATGCAAGCTGTGGGACCAAGGTGTGACGGTGTGAGCTGTGAGCCGTGATGCTCAT 1920
Qy 641 SerValIleThrTyrglnIleThrSerGlyAaThrArgAsnArgPheSerIleThrSer 660
Db 1921 AGTATCATCACTACAGATCAACAGTGGCAATATCGAAACCCCTTCTCATACACAGC 1980
Qy 661 GlnSerGlyGlyValLeuValSerLeuAlaLeuProLeuAspTyrlValLeuGluArgGln 680
Db 1981 CAAGTGTGTGGGTGTGATCCCTTGTCCCTGCACTGACATCAAACTTGAAGCGGAG 2040
Qy 681 TyrlValLeuAlaValAlaThrAlaSerAspGlyThrArgGlnPheThrAlaGlnIleValVal 700

```

---

```

Db 2041 TATGTGTGGCTGTACCGCTCCAGTGGCACTCGGACAGACAGGACAGATGTGTG 2100
Qy 701 AsnValThrAspAlaAsnThrHsArgProValPheGlnSerSerIleTyrlThrValAa 720
Db 2101 AATGTACCAAGCCCAACACCATGCTCTGTCTTCAAGCTCCACATACAGATGAT 2160
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db 2161 GTTAATGAGAGCCGCGGACAGGACCAAGTGTGTGATGACGCCACAGTGAAGAGC 2220
Qy 741 ThrGlyGluAsnAlaArgIleThrTyrlPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACAGTGAATGTCCTGCAATCACTTCAATGAGAGACAGATCCCAAGTCCGATC 2280
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrlGluAspGlnValSer 780
Db 2281 GATGCAAGACCGGGGCTGTCAACCCAGGCTGAGCTGAGCTGAGTACGAAACCAAGTGT 2340
Qy 781 TyrlThrLeuAlaIleThrAlaArgAspAspGlyIleProGlnLysSerAspThrTyrl 800
Db 2341 TACACCTGCGCATTTACTGCTGGGACAAATGAGCATTCCACAGAGTCCGACACCTAC 2400
Qy 801 LeuGluIleLeuValAspAspValAspAspAaAaProGlnPheLeuArgAspSerTyrl 820
Db 2401 CTGAGATCTGTGTGACAGCGAATGACAAATGCCCTCAAGTCTGTGAGATCTTAC 2460
Qy 821 GlnGlySerValTyrlGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGCAAGTGTATGAGATGAGATGCCACCTTCACTAGAGTCTGTGAGATTCAGCACT 2520
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrlThrPheGlnGlyGlyAspAspGly 860
Db 2521 GATCGATTCGTGACTTATGAGCAGAGGCTTCTTCACTTCCAAAGAGCCAGATGGA 2580
Qy 861 AspGlyAspPheIleValGlySerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GAGGTGACTTATTTTATGATGATCAGCTGAGCATGTGCGAAGCCTTACGAGGCTGAT 2640
Qy 881 ArgGluAsnValAlaGlnTyrlValLeuArgAlaTyrlAlaValAspLysGlyMetProPro 900
Db 2641 CGAGAAAGATGGGCCAGATGTCTTGGGGCATATGAGTGAAGAGAGAGCCCATGGGCCCA 2700
Qy 901 AlaArgThrProMetGluValThrValThrValLeuAspValAspAspAsnProProVal 920
Db 2701 GCCCGACACTTATGAGATGACATGCTGTGTGAGATGATATCAATCCCTGTGTC 2760
Qy 921 PheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTGACCAAGATGATGATTTGATGTGTGTGAAAGAGACAGCCCATTTGGGCTAGCGGTG 2820
Qy 941 AlaArgValThrAlaThrAspProAspGluGlyTyrlAsnAlaGlnIleMetTyrglnIle 960
Db 2821 GCCCGGTGACAGCCACTGACCCGATGGAAGGACCAATGCCAGTTATGTACAGATT 2880
Qy 961 ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 2881 GTGAGAGGCAATCCCTAAGGTCTTCCAGCTGGAATCTTCTCCGGAGAGCTGAGAGCC 2940
Qy 981 LeuValAspLeuAspTyrlGluAspArgProGluTyrlValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGTGATGACTTGAACAGAGAGCCGGCTGAGTATGCTCTGTGATTCAGAGGCAAGTGA 3000
Qy 1001 AlaProLeuValSerArgAlaThrValHsValArgLeuLeuAspArgAspAsnPro 1020
Db 3001 GCTCTCTGTGTGCGGGGTACAGTCAAGTCCGCTCTTCAACCGCAATGACCA 3060
Qy 1021 ProValLeuGlyAsnPheGluIleLeuPheAsnArgTyrlThrAsnArgSerSerSer 1040
Db 3061 CAGTGTGGGCACTTTATGATCTTTTCAACATATGTCAACATGTGCTGACAGC 3120
Qy 1041 PheProGlyGlyAlaIleGlyArgValProAlaHsAspProAspIleSerAspSerLeu 1060
Db 3121 TTCCCTGGGGGTGCCATTGGCGAGTACCTGCCCATGAGCCCTGATATCTCAGATAGTCTG 3180

```



QY 1061 ThrTyrSerPheGluValArgGlyValAsnGluSerLeuValLeuLeuAsnAlaSerThrGly 1080  
Db 3181 ACTTACAGCTTTGAGCGGGGAAATGAATCACTGAGCTGCTGTCATTCCTCCACAGGGT 3240  
QY 1081 GuluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
Db 3241 GAGCTGAACCTTAAGCCGCGCATGGAACACACCGGCTCTGAGGCGCTCATAGACCTG 3300  
QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIleIle 1120  
Db 3301 CTGGTCTCAGACGGCTTACACAGCTGACCGCCCACTGCGCTGCGTGTACATCATC 3360  
QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
Db 3361 ACCGATGAGATGCTCAACCCACAGCATCAGCTGCGCTGGAAGGACATGACCCGAGGCC 3420  
QY 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
Db 3421 TTCTCTGTCAACATGCTAGGCTCTTCATTCAGGCGGTGGCCGACGCTGGCCACGCCA 3480  
QY 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
Db 3481 CCGGACACAGCTGAGTGTCTTCAACGTAACAGGGGACACGAGCCGCCGAGGCGCACATC 3540  
QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhePheLeu 1200  
Db 3541 CTCAACGTAGGCTGTGGTGGGCGAGCCGCGGCGCGGCGGCGGCGGCGCTTCTC 3600  
QY 1201 ProSerGluAspLeuGlnGluValLeuIleValLeuAsnArgSerLeuLeuAlaIleSer 1220  
Db 3601 CCTCTGAGAGACTGTCAGAGAGCGCTTATACCTCAACCGCAGCTGTCAGCGCATCTCG 3660  
QY 1221 AlaGlnArgValLeuProPheAspAsnAlaCysLeuArgGluProGlyGluAsnTyr 1240  
Db 3661 GCAACAGCGGTGCTGCTTCAACAGACATGCTGCGGAGCCCTGCGAGATC 3720  
QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
Db 3721 ATGCGCTGCGTGGTGGTGGCTTCACTCGACTCCGCGCCCTTCATCCCTCTCTCC 3780  
QY 1261 ValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgAspProGlyPhe 1280  
Db 3781 GTGCTCTTCCGCGCCATCCACCCCGTGAAGGCTGCGCTGCGCGCCGCGCTTCC 3840  
QY 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProGlyProHis 1300  
Db 3841 ACGGCGTACTACGAGACCGAGGCTGACCTTGTACTCCGCGCCCTGTGGCCCCC 3900  
QY 1301 GlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
Db 3901 GGGCGCTGCGCGAGCGCGGAGGGCGGCTACCTGCTGTGTGATGACTACAGGGT 3960  
QY 1321 GlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsnGly 1340  
Db 3961 GAGCACTGTGAGGTAGTGTCTGCTCAGGCGCTGACCCCGGTGTGCAAGATGGG 4020  
QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPhe 1360  
Db 4021 GGCACCTGTGTCAACTTGTGTGGGCGGCTTCAATGGATGGATTCCTGAGACTTC 4080  
QY 1361 GlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
Db 4081 GAGAAACCTTACTGCGCAGGTCAGACGCGCAGCTTCCCGCCACTTCATCACTTT 4140  
QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
Db 4141 CGGGGCTGCGCGAGGGTGTCCACTTCACCTGGGCTCTCTGTGTGCAAAAGAGGCGC 4200  
QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluLysHisAspPheValAlaLeuGlu 1420  
Db 4201 GACGGGTGTCTGTGTATCAATGGGCGTTCATGAGAACATGACTTTGTGGCCCTCGAG 4260

QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
Db 4261 GTGATCCAGAGACAGTCCAGTCACTTCTCTGAGGGGAGTCAACCAACCAAGGTGTCC 4320  
QY 1441 ProPheValProGlyGlyValSerAspGlyGlyIleThrPheIleThrValGlnLeuLysTyr 1460  
Db 4321 CCATTGCGCCCGGAGGAGTCAGTGATGGCCAGTGGCATACGCTGACCTGAAATCAAC 4380  
QY 1461 AsnLysProLeuLeuGlnGlyIleThrGlyLeuProGlnGlyProSerGluGlnLysValAla 1480  
Db 4381 AATTAACCACTTGTGGTCAAGACAGGCTTCCACAGGCCCCATCAAGACAGAGGTGGCT 4440  
QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4441 GTGCTGACCGTGGATGGCTGTGACACAGAGTGGCTTGCGCTTGATGTGCTGGGCGC 4500  
QY 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLysSerLeuAspLeuThr 1520  
Db 4501 AACTACTCTGTGCTGCGCCAGAGGACCCAGGGGTGGACAGAAAGTCTGTGATCTGACG 4560  
QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
Db 4561 GGGCCCTGTACTAGGCGGGGTGCTGACCTTGCCGAGAGCTTCCAGTCCGAATGGCG 4620  
QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
Db 4621 CAGTTGCGGGGTGATGCTGGAACCTTGAGGTGACACCCGCGCATATGACATGGCTAC 4680  
QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysValAsnValCysAspSer 1580  
Db 4681 TTCAATGGCAACAAAGGACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4740  
QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAlaAsnGlnTyrAspAlaPheSerCysGlyCys 1600  
Db 4741 AACACTTCCACATAGGGGAGACTTGTGCTGACATGAGAGCGCTTCACTGCGAGTTC 4800  
QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisAspLeu 1620  
Db 4801 CCCCCTGCGCTTGGGGGCAAGAGCTGCGCCCAAGAAATGGCCAAATCCACAGCATTTCTG 4860  
QY 1621 GlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProThrPyrLeu 1640  
Db 4861 GCGACAGCCCTGGTGGCTGTGATGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4920  
QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCTTCATGTTCCGACCGCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4980  
QY 1661 ArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeuSerValGluGlyThr 1680  
Db 4981 CGCAGCACCATCACTCACTACCTAGAGAGGCGCAGTGTGAGCGGTGAGGGGACACA 5040  
QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThrHis 1700  
Db 5041 GGGCTTCAAGGCTCTCTCTCTGCTGTGAGCCAGGCGCGGCGCAATGAGGTACGTGGCAC 5100  
QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyLysProGlyHisAlaAlaIleLeuSerPheAsp 1720  
Db 5101 CATGCAACGTGCACTGGAGACAGCGGGGGGCTGCGCATTCGTCTGCACTTCAT 5160  
QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
Db 5161 TATGGGAG 5220  
QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
Db 5221 AACATTAACGTGGGGGAAATCTGGGCAAGCGGCGGTGTGGCCCGGCTTGTGGGCGC 5280  
QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
Db 5281 TGTTTGAGGGGTGTGGGTGAGGATACGCGCAGAGGGGTTTAAACAGCTGATCCAGC 5340  
QY 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800



Db 5341 CATGGGAGAGCAACAGTGGACCAAGCTGTAGCTGCTGACCTTGACTGACTCAAC 5400  
 Qy 1801 ProCyProAlaAsnSerTyrCySerAsnAspTrpAspSerTyrSerCySerCyAsp 1820  
 Db 5401 CCGGTCTCTGTAACAGCTATTGGACGACAGCTGGGACAGCTATCTCTGACGCTGGAT 5460  
 Qy 1821 ProGlyTyrTyrGlyAspAsnCySerThrAsnValCyAspLeuAsnProCysGluHisGln 1840  
 Db 5461 CAGAGTACTAGTGTGACAACTGACTAATGTGTGACCTGACGACCCGTGTAGACACAG 5520  
 Qy 1841 SerValCySerThrArgIleProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 Db 5521 TCTGTGTGTATCCCGCAAGCCGACGAGCCCGCCATGCTATACCTGCGAGTGTCCCGCAAT 5580  
 Qy 1861 TyrLeuGlyProTyrCysGluThrArgIleAspGlnProCysProAspGlyTrpTrpGly 1880  
 Db 5581 TACCTTGGGCGATCTGTGAGACAGAGATTACAGACTGTCTCCCGTGGCTGGTGGAGA 5640  
 Qy 1881 HisProThrCysGlyProCysAsnCyAspValSerIleGlyPheAspProAspCyAsn 1900  
 Db 5641 CATCCCACATGTGGCCCATGCACTGTGATGTGACAAAGGCTTGAACCCAGACTGCAC 5700  
 Qy 1901 LysThrSerGlyGluCysHisCysGlyGluAsnHisGlyTrpArgProProGlySerProThr 1920  
 Db 5701 AAGACAAAGCGGAGTCCCATGCAAGAGAACCACTACCGGCCCCAGGACAGCCCAAC 5760  
 Qy 1921 CysLeuLeuCyAspCysTyrProThrGlySerLeuSerArgValCyAspProGlyLys 1940  
 Db 5761 TGCTCTTGTGTATCTGTACCTGACCCAGGCTCTTGTTCAGAGTCTGTACCTTGAGAT 5820  
 Qy 1941 GlyGlnCysProCysIleProGlyValIleGlyValGlnCysAspArgCysAspAsnPro 1960  
 Db 5821 GGCAGAGTGCATGACCAAGCCAGGTGTGATCGGGGTGATGACCGCTGTGACAACT 5880  
 Qy 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle 1980  
 Db 5881 TTGTCTAGGTGACACCAACAAAGCTGTGAGTGAATTATGACACTGCCACAGACGAT 5940  
 Qy 1981 GlnAlaGlyIleTyrTrpProArgThrArgPheGlyLeuProAlaAlaIleProCysPro 2000  
 Db 5941 GAGGCTGGAGATCTGTGGCCCGGCTGACCGGCTGTGGGCTGTGCTCCCTGTGCC 6000  
 Qy 2001 LysGlySerPheGlyTyrAlaValArgHisCysAspGluHisGargIleTyrPheProPro 2020  
 Db 6001 AAGGCTCTTTTGGAGCTGTGTGCGCACCTGTGATGACGACAGGGGTGGCTCCCA 6060  
 Qy 2021 AsnLeuPheAsnCySerThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu 2040  
 Db 6061 AACCTTTCACTGACGTCATCACTCTCTCAAGAACTGAAGGCTGTGCGTGAAGGGCTA 6120  
 Qy 2041 GlnArgAsnGluSerGlyLeuAspSerGlyArgSerGlnGluLeuAlaLeuLeuArg 2060  
 Db 6121 CAGCGAATGAGTCAAGGCTTGACTCAAGGGGCTCCAGACGACGATGCTGTCTGCGC 6180  
 Qy 2061 AsnAlaThrGlnHisGthrAlaGlyTyrPheGlySerAspValLysValAlaTyrGlnLeu 2080  
 Db 6181 AACCCACGCGACGACACAGCTGTACTTGTGCGCACGTCGCAAGGTGGCTCTCAAGCTG 6240  
 Qy 2081 AlaThrArgLeuLeuAlaHisGluSerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 Db 6241 GCCACGGGCTGTGGCCACGAGACCCAGCGGGGCTTTGGGCTGTGTGCCACACAG 6300  
 Qy 2101 AspValHisPheThrGluAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
 Db 6301 GACGTGACACTTCACTGAGAACTCTGCTGGGTGGGACGCGCTCTCGACACAGCCAAC 6360  
 Qy 2121 LysArgHisTrpGluLeuIleGlnGlnThrGluGlyGlyTyrAlaAlaTrpLeuLeuGlnHis 2140  
 Db 6361 AAGGGGCACTGGGAGTGTATCAGACAGAGGGGTGGCACCGCTGGCTGTCCACAC 6420  
 Qy 2141 TyrIleAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160

Db 6421 TATAGGCTTACGCCAGTGGCTTGGCCGACGAACATGGGACACCTTACTTAAGCCCTTC 6480  
 Qy 2161 ThrIleValThrProAsnIleValIleSerValValArgLeuAspIleGlyAsnPheAla 2180  
 Db 6481 ACCATGTGACCGCCCAACATTTGTCTATCTCCGTAGTGGCTTGACAAAGGAACTTGTCT 6540  
 Qy 2181 GlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGlyGlnProProAspLeuGluThr 2200  
 Db 6541 GGGGCAAGCTGCGCCGCTTACAGGCGCTGCTGGGAGAACACCCCGGACCTTGAGACA 6600  
 Qy 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
 Db 6601 ACAGTATTTCTGTGAGTGTGTCTTCAAGAGAGCCCGCTGTGTGAGGCCGAGGC 6660  
 Qy 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGlnArgArgHisAspProGluLeu 2240  
 Db 6661 CCGGAGAGGCCGAGAGGACAGAGAGCTGGGACCGGCGACAGAGCCGAGCGAGCTGG 6720  
 Qy 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 6721 AGCCAGGGTGAAGGCTGTGGCCAGCGCTCATCTACCTGACCCGCTGCTACTGCTCT 6780  
 Qy 2261 HisAsnTyrAspProAspLysArgSerIleuArgValProLysArgProIleIleAsnThr 2280  
 Db 6781 CATACTATGACCTTGACAGCGCAGCTTGAGAGTCCCAACGCCGATCATCAACACA 6840  
 Qy 2281 ProValIleSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLys 2300  
 Db 6841 CCGGTGTAGGATCAGCTGCTCATGATGATGAGAGGCTTCTGCCCCGGGCTCGACAA 6900  
 Qy 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrLysProIleCysVal 2320  
 Db 6901 CCGGTCAAGGTGACAGTTCGCTGCGAGAGACAGAGAACGACCAAGCCATCTGTGTGC 6960  
 Qy 2321 PheTrpAsnHisSerIleLeuValSerGlyThrGlyIleTrpSerAlaArgIleCysGlu 2340  
 Db 6961 TTCTGACCACTTCAATCTGTGTCAAGTGCACAGGTGTGTGTGGCCAGAGGCTGTGA 7020  
 Qy 2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
 Db 7021 GTGCTTCCGCAATGAGAGCCACGTCAGCTGCAAGTCAACATGACAGACTTGCTGCT 7080  
 Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThr 2380  
 Db 7081 GTGCTATGAGAGCTTCTGTGGGAGAGTGGAGATCTCTCCATGAAGACATCGACA 7140  
 Qy 2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu 2400  
 Db 7141 TACGTGGCTGTAGGTACCTTGGCTGCTCTTGTCTCACTCTTCTTCTCACTCTCTC 7200  
 Qy 2401 LeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
 Db 7201 TTGCGTATCTGCGCTCCCAACCAACAGGCACTCCGATCCGACCTGCGCTGCGGC 7260  
 Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
 Db 7261 CTGGCTCAGCTGTCTTCTCTGTGGAGATCAACAGGCTGACCTCCCTTTTGTGCA 7320  
 Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460  
 Db 7321 GTATATTGCATCTGTGTGACCTTCCGTATCCTGTGACCTTTTCCGGGCTGTGCGAG 7380  
 Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProPheCysArg 2480  
 Db 7381 GCCTTGCACCTGTACCGGCGACCTCACTGAGTGGCGGAGTCAACACCGGCCCAAGCCG 7440  
 Qy 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
 Db 7441 TTCTACTACATGTGTGGCTTGGGCGTGTGCTGTCTGTATCAAGGGCTTACCGTGGCTG 7500  
 Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThrLeuIle 2520  
 Db 7501 GACCCGAGGGGCGAGGAAACCTGACTTGTGCTGTCTCATCTATGACAGGCTGATC 7560

QY 2521 TTPSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
 DB 7561 TGGAGTTTTCGCGGCCGGGTGGCTTTCGCTTCGATGAGTGTCTTCTTCATCAATCCG 7620  
 QY 2541 AAlaAlaArgAlaSerCysAlaAlaGlnAlaGlnGlyPheGlnLysValGlyProValSer 2560  
 DB 7621 GGGGCCCCGGGCTCTCGTGTCTGCGCGCGGCTTTCGAGAAAGAGTCTCTGCTG 7680  
 QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeu 2580  
 DB 7681 GGCTTCGACCCCTCTTCGCGCTTCCTGCTGCTGAGCGCACTGCGCTGCGACCTG 7740  
 QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIle 2600  
 DB 7741 CTCTGCTCAACAGGACACCCCTCCCTCACTACCTTTGGCTACCTCGCAATTGCATC 7800  
 QY 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGlnValLysValAlaLeu 2620  
 DB 7801 CAGGGGCCCTTCATCTTCTCTCTATGTGTGCTTTCAGCAAGAGAGTCCGGAAGCACTC 7860  
 QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrTrpSerThrLeu 2640  
 DB 7861 AAGCTTGCTGAGCGCGCAAGCCCTGACCTGCTGACCAACAGTCCACCTG 7920  
 QY 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
 DB 7921 ACCTGCTTCAACATCGCCCGACCCCTACGCAAGTGGCGGCTGTACCAAGCTTACGGA 7980  
 QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
 DB 7981 GACTGGCGCGGCTCTCTGCAACGACCACTCTCCGCGCAAGATCAGCCGCACTACATC 8040  
 QY 2681 ProPheLeuLeuArgGlnLysSerAlaLeuAsnProGlnGlnGlyProProGlyLeuGly 2700  
 DB 8041 CCTTCTTCTGCTAAGGAGAGTCCGCACTGAAACCTGCGCAAGGCGCCCTGCGGCGG 8100  
 QY 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
 DB 8101 GATCCAGGACGCTGTCTCTGGAAGGTCAAGACCAAGACAGATCTGACACGGAATCC 8160  
 QY 2721 AspSerAspLeuSerLeuGlnLysAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
 DB 8161 GACAGTGAACCTGTCTCTTACAAACAGCAAGATGGCTCTTACCTTACCACTCACTCA 8220  
 QY 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 DB 8221 GACAGTGAAG 8280  
 QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGlnArgLeuProLeuHisSerThrProLysAsp 2780  
 DB 8281 TGGGATAGCTGCTGGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8340  
 QY 2781 GlnGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGln 2800  
 DB 8341 GGGGGGCGGAGGCTGTGGCAAGGCGCCCTGCGCAGAGACTTGTGGACCAACAAAGAG 8400  
 QY 2801 SerSerGlyAsnGlyAlaProGlnGlnLysArgLeuArgGlnLysAsnGlyAspAlaLeuSerArg 2820  
 DB 8401 AGTATGGCAAGGGGCGCCCTTACAGAGCGGCTGCGGAGAAATGAGAAATCCCTGCTGCA 8460  
 QY 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLys----- 2836  
 DB 8461 GAGGGGTCCTCTAGGCGCCCTTCAAGGCTCTTCTGCGCAAGCTCACAAAGGTAGTGGGCG 8520  
 QY 2836 ----- 2836  
 DB 8521 ACCCCCACTGCGAGCTCCCTAGTACAGAGCTCATACCTCATCTCTGTGGCGG 8580  
 QY 2837 -----GlyLeuLeuLysCysLeuProThrIle 2847  
 DB 8581 CACCTCAAGCCCGCGCGCGCCACAGGATCTTACAGAAAGTGTCTGCCCACTC 8640

QY 2848 SerGlySerSerLeuLeuArgLeuProLeuGlnGlnCysThrGlySerSerArgGly 2867  
 DB 8641 AGCGAAGAGAGAGCTCTTCGCTCCCTCGAGCAATGACAGGATCTTCCCGGGGC 8700  
 QY 2868 SerSerAlaSerGlnGlySerArgGlyGlyProProPheArgProProProArgGlnSer 2887  
 DB 8701 TCTTCGCTAGTGAAGGAGCGCGGGGCGGCGCCCTCCCGCCACCGCGGAGAGC 8760  
 QY 2888 LeuGlnGlnGlnLeuAsnGlyValMetProIleAlaMetSerIleLysAlaGlyThrVal 2907  
 DB 8761 CTCAGAGAGAGCTGAGAGGAGGTATGCTCCATGCGCATGAGCATCAAGGCAAGGTG 8820  
 QY 2908 AspGluAspSerSerGlySerGluPheLeuPhePheAsnPheLeuHis 2923  
 DB 8821 GATGAGACTCGTACAGGCTCCGAATTCTCTTCTTAATCTTCCTGAT 8868  
 RESULT 4  
 ABX34546  
 ID ABX34546 standard; cDNA; 11762 BP.  
 AC ABX34546;  
 DT 13-FEB-2003 (first entry)  
 XX  
 DE Human mdt cDNA SEQ ID 107.  
 XX  
 KW MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotoxic; antianemic; antiproliferative; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;  
 KW ananemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis; gene; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 29-MAR-2001; 2001US-280068P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 19-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JT, Tusson O, Yip PP, Ambrey SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urabaka MB;  
 XX  
 DR WPI; 2003-058431/05.  
 DR P-PSDB; ABU11556.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis  
 XX  
 PS Claim 1; SEQ ID NO 107; 339pp + Sequence Listing; English.  
 CC This invention describes a novel disease detection and treatment molecule  
 polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,

CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotoxic,  
 CC antianemic, antipruritic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndrome, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
 CC hepatitis. AX34440-ABX34835 encode the MDR1 polypeptides represented in  
 CC AB011450-AB011845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 11762 BP; 2286 A; 3738 C; 3283 G; 2455 T; 0 other;

# Alignment Scores:

Pred. No.:	0	Length:	11762
Score:	15393.00	Matches:	2894
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	59.02%	Indels:	0
DB:	25	Gaps:	0

US-09-916-849A-3 (1-2923) x ABX34546 (1-11762)

QY 30 LeuGIyAaPpGluValIyGlyProCyAaRgSerLeuGIySerAaRgIyAaRgIySerSerGIy 49  
 DB 2 TTGGAGAGCCAAAGTGGGGCCCTGTCGTCCTGGGGGTCAGAGGAGCGAGGCTCTTGGGG 61  
 QY 50 ALaCyAlaPAPrometGIyTyrLeuCyAaRgSerSerAaSerAaSerLeuTyrLeuTyrThr 69  
 DB 62 GCTGCGGCCCCAATGGGCTGCTGCTGCTCACTTCAAGGCTGGAACCTTGGCTCTTCAACC 121  
 QY 70 SerAaRgCyAaRgAaPAlaGIyThrGIyLeuThrGIyHISLeuValProHISaAaPGLy 89  
 DB 122 AGCCGCTGCGAGGAGTGGCGGCACTGAAGCTGACCTGACCTGATGCCACACAGATGGC 181  
 QY 90 LeuAaRgValTyrCyAaRgProGIySerGIyAaHISaIleProLeuProProAlaProGIyGIy 109  
 DB 182 CTGAGGATTTGGTTCAGAAATCCAGAGCCCAATATCCCTTACCAACAGCTCTGAAAGGC 241  
 QY 110 CyAaProTyrSerCyAaRgLeuLeuGIyTLeGIyGIyHISLeuSerProGIyGIyLeu 129  
 DB 242 TGCCCTCGAAGCTGCTGCTCTCTGCGCATTTGAGGCGCACTTTCCCAAGGAGGAGCTC 301  
 QY 130 ThrLeuProGIyLuhISProCyAaLeuLysAaProAaRgLeuAaRgCyAaGIySerCyAaLys 149  
 DB 302 ACACGCGCGAGAGACCCGCTTAAAGGCTCCACCGGCTCAGATGCCAATGCCCTGCAAG 361  
 QY 150 LeuAaGIaIaAProGIyLeuAaRgAlaGIyGIyAaRgSerProGIyLuhISLeuGIy 169  
 DB 362 CTGGACAGAGGCCCCGGGCTCAGGCGAGGAGAAAGTCCACAGAAAGATCCCTGGAGTGG 421  
 QY 170 ArgAaRgIyAaRgAaValAaHISaIleProGIyAaHISaIleProProSerTyrGIyAaIaThr 189  
 DB 422 CGTCGAAAGAAATGTAATACAGCCCCCAAGCTTCCAGGCCCCCAAGCTTACAGGCGCA 481  
 QY 190 ValProGIyLuhISProAlaGIyThrProValAaIaSerLeuAaRgAlaIleAaProAaP 209  
 DB 482 GTGCGGAGAAACCAACCAAGAGCAACCCCTGTTCATCCCTGAGGCGCATGAGCCCGGAC 541  
 QY 210 GIuGIyGIuAaGIyAaRgLeuGIyTyrThrAaPAlaLeuPheAaPpSerAaRgSerAaP 229  
 DB 542 GAGGATGAGGCGAGCTGAGTGAATACATGATGATCCCTTTGATAGCCGCTCCAC 601  
 QY 230 GluPhePheSerLeuAaPProValAaRgIyAaIaIaThrAaIaGIyLuhISLeuAaPArg 249  
 DB 602 CAGTCTTCTCCCTGAGCAAGTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 661  
 QY 250 GluThrIySerThrHISaIleAaPheAaRgValAaIaGIyAaHISaIleGIyMePProAaRg 269

DB 662 GAGACCAAGAGAGCCACCGCTTCAAGGTCAAGGCGAGAGACCAAGGCAATGCCCGACCA 721  
 QY 270 SerAaIaLeuAaIaThrLeuThrIleLeuValAaThrAaPpThrAaAaPpProValAaPhe 289  
 DB 722 AGTGGCTGGCTGATCACTCAACATCTGGTTAATGAGCAACAAATGACATGACCTGAGTTC 781  
 QY 290 GIuGIyGIuGIyTyrIyGIySerLeuAaRgGIyAaHISaIleGIyTyrGIyValLeu 309  
 DB 782 GACACGAGAGATGACAGAGAGCTTCAAGGAGAACTTGAAGGTTGGCTTATAGAGTCTC 841  
 QY 310 ThrValAaRgAlaThrAaPpGIyAaPAlaProProAaAaIaAaIleuTyrAaRgLeu 329  
 DB 842 ACTGACAGGCGCAACGATGGTGAATGCCCTCCCAATGCAATATTCATGATCCGCTGCTG 901  
 QY 330 GIuGIySerGIyGIySerProSerGIyValAaPheGIyAaAaProAaRgSerGIyValIle 349  
 DB 902 GAGGAGCTTGGGGGAGGCCCCCTTGAAGTCTTGAATGACCCCTGCTTGGGGGATC 961  
 QY 350 ArgThrAaRgIyProValAaPpAaRgGIyValGIySerTyrGIyLeuThrValGIyAaIa 369  
 DB 962 CGAACCCTGAGCTGCTGATGATGAGAGAGTGAATCTTACAGCTGACGATGAGAGCA 1021  
 QY 370 SerAaPpGIyAaRgAaPProGIyProAaRgSerThrAaIaAaIaPheLeuSerVal 389  
 DB 1022 AGTGAACAGAGGTGGGACCGGCTCTCGAGATACCAAGCCGCTGTTCTTCTGCTG 1081  
 QY 390 GluAaPpAaAaPpAaAaPpAaPProGIyAaHISaIleAaRgTyrValAaGIyValAaRg 409  
 DB 1082 GAGATGATCAATATATATGCCCCAGTTTGAAGAGGCTATATGATGATGATGATGATGATG 1141  
 QY 410 GluAaPpValThrProGIyAaPProValAaRgValAaThrAaSerAaPpAaRgAaPpGIy 429  
 DB 1142 GAGGATGTGATCTCAAGGAGGCCCCAGTATCTCGAGTCAAGCTCGAGTGAAGCAAGGGG 1201  
 QY 430 SerAaAaIaValAaHISaIleTyrSerIleMetSerGIyAaAaIaAaRgIyGIyIlePheTyrLeu 449  
 DB 1202 AGCAATGCCCTGTGTGATATATGATCAATGATGAGGAAATGCTGGGAGCAAGTTTATCTG 1261  
 QY 450 AaPAlaGIyThrGIyAaIaLeuAaPValAaIaSerProLeuAaPpTyrGIyThrIyRgIy 469  
 DB 1262 GATGCCAGAGCTGAGCTCTGATGTGTGAGGCTCTTCACTATGATGATGATGATGATGATG 1321  
 QY 470 TyrThrLeuAaRgValAaRgAaIaAaPpGIyGIyAaRgProProLeuSerAaValAaSerGIy 489  
 DB 1322 TACACCTTACGGGTGAGAGCAAGATGGTGGCCGCCCTCCCACTCTTATATGCTCTGCGC 1381  
 QY 490 LeuValThrValGIyValLeuAaPpIleAaAaPpAaAaIaProIlePheValAaSerThrPro 509  
 DB 1382 TTGGTGAAGTATGAGGTCTGATATCAACAGAAATGCCCTTCCTTCGACACACCCCT 1441  
 QY 510 PheGIyAaIaThrValIleuGIySerValProLeuGIyTyrLeuValAaHISaIleGIyAaIa 529  
 DB 1442 TTCCAGGCTATCTGCTCTGAGAGAGCTGCCCTTATGCTATGCTTCTTCAATGTCAGGCT 1501  
 QY 530 IleAaPAlaAaPAlaGIyAaAaAaAaIaAaRgLeuGIyTyrAaRgLeuAaGIyAaIaHIS 549  
 DB 1502 ATGCAAGCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1561  
 QY 550 AaPpPheProPheThrIleAaAaAaGIyTyrGIyTyrIleSerValAaIaGIyLuhISLeuAaP 569  
 DB 1562 GACTTCCCTTCAACATCAACATGAGGAGGCTGATGATGATGATGATGATGATGATGATG 1621  
 QY 570 ArgGIyGIyValAaPpPheTyrSerPheGIyValGIyAaIaAaRgAaPpIleGIyThrProAla 589  
 DB 1622 CGGAGGAAATGATATTTCTATCACTTTGGGTGAGAGCTGAGACATGAGCACTCCAGCA 1681  
 QY 590 LeuThrAaSerAaSerValAaSerValAaIaLeuAaPpAaIaAaAaAaAaPProThr 609  
 DB 1682 CTCACCTGCTCGGCGAGTGTCAAGTCACTGATGATGATGATGATGATGATGATGATGATG 1741  
 QY 610 PheThrGIyProGIyTyrThrValAaRgLeuAaGIyAaAaIaAaIaValAaGIyThrSerVal 629

Dd	1742	ITTAACCAACAGAGTACACAGTCGGCGCTCAATGAGATGCACTGTGGGACCAAGCGTG	1801
Qy	630	ValThrValISerAlaValAspArgAspAlaHisSerValIleThrTyGlnIleThSer	649
Dd	1802	GTGACGGGTGTCAGCTGTGGACCGTGATGTGTCAATATGTATCACTTACAGATCACAGT	1861
Qy	650	GlyAenThrAArgAsnArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeu	669
Dd	1862	GGCATATCTCCGAACCCGCTTCTCCATTCACAGCCAAAGTGGTGGGCTGGTATCCCTT	1921
Qy	670	AlaLeuProLeuAspTyrlTybLeuGlnIleuArgGlnTyrrValIleuAlaValThrAlaSerAsp	689
Dd	1922	GCCCTGCCACCTGGACTTACAAATTAAGCGGGCAGTATGGTGGCTGTATCCGCCCTCCAT	1981
Qy	690	GlyThrArgGlnAspThrAlaGlnIleValValaenValIThrAspAlaAsnThrHisArg	709
Dd	1982	GGCATCGGGCAGGAGACCGGCACAGTTGTGGATGTCCACCGAGCCACACCATGT	2041
Qy	710	ProValPheGlnSerSerHisTyrrThrValaenValaenGluuAspArgProAlaGlyThr	729
Dd	2042	CCTGCTCTTCAGAGCTCCACCACTATCACTGATGTATTATGAGAGACCGGCCGAGGACAC	2101
Qy	730	ThrValValIleuIleSerAlaThrArgGluAspThrGlyGlyuAsnAlaArgIleThrTyrr	749
Dd	2102	ACGGTGGTGGTCATCAGGCCACCGGATAGAGACACAGGTGAGATGCCCGATCACCTAC	2161
Qy	750	PheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThr	769
Dd	2162	TTCATGAGAGGACAGCATTCGCCCATGTCCGATTCAGATGCAGACCGGGGGCTGTCAACCC	2221
Qy	770	GlnAlaGluLeuAspTyrrGluAspGlnValSerTyrrThrLeuAlaIleThrAlaArgAsp	789
Dd	2222	CAGCGTGAAGTCGACCTACGAGAACCAAGTGTCTTACCCCTGGCCATTACTGTCCGGGAC	2281
Qy	790	AsnGlyIleProGlnIleuSerAspThrThrTyrrLeuGlnIleLeuValaenAspValaen	809
Dd	2282	AATGGCATTTCCCAAGAAAGTCCGACCAACCACTACCGAGGAATCCGTGTCAAGAGTGTGAT	2341
Qy	810	AspAsnAlaProGlnPheLeuArgAspSerTyrrGlnGlySerValTyrrGluAspValPro	829
Dd	2342	GACAAATGCCCTCACTTCTCTCGAGACTTCTTACCGAGGCAATGTCTTATGAGAGATGCCA	2401
Qy	830	ProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArg	849
Dd	2402	CCCTTCACTACGTCGCTCTGCAAGATCTCAACCACTGATCGTATTCGTGACTTAATGGCAAG	2461
Qy	850	ValPheTyrrThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValGluSerThr	869
Dd	2462	GTCTTCTTACACCTTCCAGAGGAGCGCACATGTAGACGCTGACTTATTTGTGTAGCTCCAGC	2521
Qy	870	SerGlyIleValArgThrLeuArgArgLeuAspArgGluuAsnValaGlnTyrrValLeu	889
Dd	2522	TCAAGGATCGTGGCGAACCGCTACGAGGCTGTGATCGAGAGAACGTCGCCACTATATGCTTG	2581
Qy	890	ArgAlaTyrrAlaValAspPheGlyPheProProAlaArgThrPrometGluValThrVal	909
Dd	2582	CGGGCATATGACATGCGACAGAGGGAGTCCCCCAAGCCGCGCACACTATGAGATGACAGTC	2641
Qy	910	ThrValLeuAspValaenAspAsnProProValPheGlnGluAspGluPheAspValPhe	929
Dd	2642	ACTGTGTGTGATGTGAATGACAAATCCCCCTCTTCTTGTGACGAGTATGAGTTGATGTGTTT	2701
Qy	930	ValGlnGluuAsnSerProIleGlyLeuAsnAlaValaArgValIThrAlaThrAspProAsp	949
Dd	2702	GTGAGAGGAACAGCCCAATTTGGGCTTACCGTGGCCCGGGGTCAAGCCATCGAACCCGAT	2761
Qy	950	GluGlyIThrAsnAlaGlnIleMetTyrrGlnIleValGlnGlyAsnIleProGluValaPhe	969
Dd	2762	GAAAGGCAACAAATGCCCAATTAATGACAAATTTGTGAGGGGCAACATCCCTGAAGTCTTC	2821
Qy	970	GlnLeuAspIlePheSerGlyGlyLeuThrAlaLeuValaenAspTyrrGluAspArg	989
Dd	2822	CAGCTGACATCTTCTCCGGGGAGCTGACAGCCCTCGTATGACTTATGACTACGAGAACCGG	2881

QY	990	ProGluTYrValLeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrVal	1009
Db	2882	CTGAGTACGCTCTGGTATCATCAGGCCACGTCACTCTCTGGTGAAGCCGGGCTACAGTC	2941
QY	1010	HisValAlaGluLeuAspArgAsnAspAsnProProValLeuGlyAsnPheGluIleLeu	1029
Db	2942	CACGTCCGCTCTTGAACCGCAATGACACCAACCAACGCTCGGGCACTTTGAATCTT	3001
QY	1030	PheAsnAsnTYrValThrAsnArgSerSerPheProGlyValIleGlyArgVal	1049
Db	3002	TTCAACAACATATGTCACCAATCGCTCAAGCAGCTCCCTGGGGGTGCATTTGGCCGAGTA	3061
QY	1050	ProAlaHisAspProAspIleSerAspSerLeuThrTYrSerPheGluArgGlyAsnGlu	1069
Db	3062	CTGCGCCATGACCTCGATATCTCAATATGTTGACTTACAGCTTTAGACGGGAAATGAA	3121
QY	1070	LeuSerLeuValLeuLeuAsnAlaSerThrGlyIleuLeuYsLeuSerArgAlaLeuAsp	1089
Db	3122	CTCAGCTGTGCTGTCTCAATGCTCCACCGGGTGAAGCTGAAGCCCGGCACTGGAC	3181
QY	1090	AsnAsnArgProLeuGluAlaIleMetSerValLeuValSerAspGlyValHisSerVal	1109
Db	3182	AACAACCGGCTCTGAGGCGCATATAGCGCTGTGTGTCAAGCGCGTACACAGCGTG	3241
QY	1110	ThrAlaGlnCysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIle	1129
Db	3242	ACCGCCCACTGCGGGCTCGTGTACATCACTACCAATGATGCTCAACCCACAGATC	3301
QY	1130	ThrLeuArgLeuGluAspMetSerProGluArgPheLeuSerProLeuLeuGlyLeuPhe	1149
Db	3302	ACGCTGCGCTCTGAGAGACATGTCAACCGAGGGCTTCTGTCAACAATGATAGGCTCTTC	3361
QY	1150	IleGlnAlaValAlaAlaThrLeuAlaIleThrProProAspHisValValIlePheAsnVal	1169
Db	3362	ATCCAGGGGTGGCGGCAAGCTGGCCACCGCACCGGACCAACGGTGTCTTCAACGTA	3421
QY	1170	GlnArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGln	1189
Db	3422	CAGCGGAGCACCGACGCCGCCCGGGGGGCACATCTCAACATGAGCTGTGGTGGGCGAG	3481
QY	1190	ProProGluYrProGlyGlyYrProProPheLeuProSerGluAspLeuGlnGluArgLeu	1209
Db	3482	CCGGCAGGGCCCGGGGGCGGGCGGCTCTCTCGCTCTAGGACCTGCAGAGGCCCTTA	3541
QY	1210	TYrLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValLeuProPheAspAsp	1229
Db	3542	TACCTCAACCGCAGCTCTGACGGCCATCTCGGCACAGCGCGTCTGCCCTTGACGAC	3601
QY	1230	AsnIleCysLeuArgGluProCysGluAsnTYrMetArgCysValSerValLeuArgPhe	1249
Db	3602	AACATCTGCTCGCGGAGCCCTCGGAAATCAATCAATGCGTGGTGGTGGCGCTTC	3661
QY	1250	AspSerSerAlaProPheIleAlaSerSerSerValLeuPheArgProIleHisProVal	1269
Db	3662	GACTCTCTCGGCGCTTCATGCGCTCTCTCTCGTGTCTTTCGGGCCATCAACCCGCTC	3721
QY	1270	GlyGlyLeuArgCysArgCysProProGlyIlePheThrGlyAspTYrCysGluThrGluVal	1289
Db	3722	GGAAGGGCTGGGCTGCGCTGCGGCCCGGCTTACCGGCTACATCTGCGAGACCGAGGTG	3781
QY	1290	AspLeuCYrTYrSerArgProCysGlyProHisGlyArgCysArgSerArgGluGlyGly	1309
Db	3782	GACCTCTGCTACTGCGCGGCTGTGGCCCCACCGGGGCTCGCGCACCGCAGGGGGCGC	3841
QY	1310	TYrThrCysLeuCysArgAspGlyTYrThrGlyGlyHisIleCysGluValSerAlaArgSer	1329
Db	3842	TACACCTGCGCTCTGTCTGATGTGTACACGGGTGAGACTGTGAGGTGATGCTCGCTCA	3901
QY	1330	GlyArgCysThrProGlyValCysValAsnGlyGlyThrCysValAsnLeuLeuValGly	1349
Db	3902	GGCGGTGGCACCCCGGGGTGTCTGAGAAATGGGGGCGACTGTGTCAACCTGCTGTGGCG	3961

QY 1350 GlyPheLeuCyAspCyProSerGlyAspPheGluLeuSerProTyCySerGlnValThrThr 1369  
Db 3962 GGTTTCAGTGGGATTTGGCCATCTGGAGACTTGGAGAGGCCCTCACTGCCAGGTGACCAAG 4021  
QY 1370 ArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPhe 1389  
Db 4022 CGCAGCTCTCCGCCCACTCTTCATCATCCTTTCGGGGCTGGCCAGCGCTTTCACCTTC 4081  
QY 1390 ThrLeuAlaLeuSerPheAlaThrLySGluArgAspGlyLeuLeuLeuTyArgAsnGlyArg 1409  
Db 4082 ACCCTGGCCCTCTCGTTTGGCAAGAGAGCGGAGGTGCTGTGTATCAATGGGCGGT 4141  
QY 1410 PheAsnGlnLysHisAspPheValAlaLeuGlnValIleGlnGlnGlnValGlnLeuThr 1429  
Db 4142 TTCAATGAGAGATGACTTGTGTGGCTTCGAGGTGATCCAGAGAGAGGTCCAGCTCAAC 4201  
QY 1430 PheSerAlaGlyLeuSerThrThrThrValSerProPheValProGlyGlyValSerAsp 1449  
Db 4202 TTCTTCGACAGGGAGGACACACACCGGTGTCCCATTCGTGCTCCGAGGAGTCACTGAT 4261  
QY 1450 GlyGlnTrpHisThrValGlnLeuLysTyrrTyrrAsnLysProLeuLeuGlnThrGly 1469  
Db 4262 GGCCTAGTGGCATACGGTGCAGCTGAATTAATCAATAGCCACTGTGGGTCAAGACAGG 4321  
QY 1470 LeuProGlnGlyProSerGlnGlnLysValAlaValThrValAspGlyCyAspThr 1489  
Db 4322 CTCCTCAAGGGGCCCATCAAGAGAGGTGGCTGTGTGATCGTGAATGGCTGTGACACA 4381  
QY 1490 GlyValAlaLeuArgPheGlySerValLeuGlyAsnTyrrSerTyrrAlaIleGlnGlyThr 1509  
Db 4382 GAGGTGGCTTGGCTTGGATGTGTCTGGGCACTACCTCTGTCTGCCAGGGCACCC 4441  
QY 1510 GlnGlyGlySerLysLysSerLeuAspLeuThrGlyProLeuLeuLeuGlnGlyValPro 1529  
Db 4442 CAGGTGGACAGCAAGAGTCTTGATCTGAACGGGGCCCTGCTACTAGCGGGGTGCTCT 4501  
QY 1530 AspLeuProGlnSerPheProValArgMetArgGlnPheValGlyCyMetArgAsnLeu 1549  
Db 4502 GACCTCCCGGAGAGCTTCCCACTGCAATGCGGCACTCTGTGGCTGACATGCGAAGCTTG 4561  
QY 1550 GlnValAspSerTrpHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrValPro 1569  
Db 4562 CAGGTGACAGCGGGCACTAGACATGGCTGACTTCAATGCAACATGGACACCGTGGCT 4621  
QY 1570 GlyCyProAlaLysLysAsnValCyAspSerAsnThrCyHisAsnGlyGlyThrCys 1589  
Db 4622 GGTGCTCGTCCAGAGAGAACTGTGTGACAGACACTTGGCAATGGGGCACTTGC 4681  
QY 1590 ValAsnGlnTrpAspAlaPheSerCySerGluCyProLeuGlyPheGlyGlyLysSerCys 1609  
Db 4682 GTGACCAAGTGGGACCGTTCAGCTGCAAGTGCCTTGGGCTTGGGGCAAGACTGC 4741  
QY 1610 AlaGlnGlnMetAlaAsnProGlnHisPheLeuGlySerSerLeuValAlaTrpHisGly 1629  
Db 4742 GCCCAGAGAAATGGCAATCCACAGCACTTCTGGGACAGCTGTGGGTGGCAATGGC 4801  
QY 1630 LeuSerLeuProIleSerGlnProTyrrTyrrLeuSerLeuMetPheArgThrArgGlnAla 1649  
Db 4802 CTCTCGCTGCCATCTCCCAACCTGTACTCAAGCTCATGTTCTGGCAAGGCCCAAGGCC 4861  
QY 1650 AspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArg 1669  
Db 4862 GAGCGGTCTCTGTGACGGCATCAACAGGGGGCGGCAACATCAACCTCAACGCTACCA 4921  
QY 1670 GlnGlyHisIleValMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeu 1689  
Db 4922 GAGGGCCACTGATGCTGACGCTGGAGGGGCAAGGGCTTCAAGGCTTCTCTCCGCTCG 4981  
QY 1690 GlnProGlyValArgAlaAsnAspGlyAspTrpHisHisAlaGlnLeuAlaLeuGlyAlaSer 1709  
Db 4982 GAGCCAGGCGGGGCAATGACGGTGACTGGCAACATGCAAGCTGGGCACTGGAGCCAGG 5041  
QY 1710 GlyGlyProGlyHisAlaIleLeuSerPheAspTyrrGlyGlnGlnArgAlaGlnGlyAsn 1729

Db 5042 GGGGGGCGCTGGCCATGGCATTTCTGTCTTCATTAATGGGCAAGAGAGAGGGCAAC 5101  
QY 1730 LeuGlyProArgLeuHisGlyLeuHisIleLeuSerAsnIleThrValGlyGlyIleProGly 1749  
Db 5102 CTGGGCGCCCGCTGCATGCTGTGACCTTGACCAATAGCAAGTGGGGGAATACCTGGG 5161  
QY 1750 ProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAsp 1769  
Db 5162 CAGCGCGGCTGTGGCCCTTGCTTGGGGCTTTTGGAGGGTGTGGGTGAGCAT 5221  
QY 1770 ThrProGlnGlyValAsnSerLeuAspProSerHisGlyGlnSerIleAsnValGlnGln 1789  
Db 5222 AGCCAGAGGGGGTTAAACAGCTGATCCAGCCATGGGAGAGATCAACGTGAGCA 5281  
QY 1790 GlyCySerLeuProAspProCyAspSerAsnProCyProAlaAsnSerTyrrCySer 1809  
Db 5282 GGTGTAGCTGTGCTGACCTTGTGATCAACCCGTGTCTGTCAACAGCTATTTGACAGC 5341  
QY 1810 AsnAspTrpAspSerTyrrSerCySerCyAspProGlyTyrrTyrrGlyAspAsnCyThr 1829  
Db 5342 AACGACTGGGACAGCTATTTCTGACGCTGTATCCAGGTTACTATGTGACACTGTACT 5401  
QY 1830 AsnValCyAspLeuAsnProCysGlnHisGlnSerValCyThrArgLysProSerAla 1849  
Db 5402 AATGTGTGACCTGAACCCGTGTGAGCACAGTCTGTGTATCCCGCAAGCCAGTGC 5461  
QY 1850 ProHisGlyTyrrThrCyGlnCySerProAsnTyrrLeuGlyProTyrrCyGlnLysArg 1869  
Db 5462 CCCCATGGCTAATCTGAGAGGTCTCCCAATTAATCTTGGGCACTAGTGAAGACAGG 5521  
QY 1870 IleArgGlnProCysProArgGlyTyrrTrpGlyHisProThrCyGlyProCysAsnCys 1889  
Db 5522 AATTGACAGCTTGTCTTGTGCTGGGTGAGGAGCATCCCATATGTGGCCATGACACTGT 5581  
QY 1890 AspValSerLysGlyPheAspProAspCyAsnLysThrSerGlyGlnCysHisCysLys 1909  
Db 5582 GATGTCAACAAAGGCTTGAACCCAGACATGCAACAGAGGGGAGGCACTGCAAG 5641  
QY 1910 GlnAsnHisGlyTyrrArgProGlnGlySerProThrCysLeuLeuCyAspCyThrProThr 1929  
Db 5642 GAGAACCACTACCGGCCCCCAGGACAGCCCACTGCTCTTGTGTGACTGTAACCCCA 5701  
QY 1930 GlySerLeuSerArgValCyAspProGlnAspGlyGlnCysProCysLysProGlyVal 1949  
Db 5702 GGTCTCTTGTCCAGATGCTGTGAACCTGAGATGGCAATGCCAAGCAAGCAAGTGC 5761  
QY 1950 IleGlyArgGlnCysAspArgCyAspAsnProPheAlaGlnValThrThrAsnGlyCys 1969  
Db 5762 ATCGGGCGTCAGTGTGACCGCTGTGACAACTTGTGTGAGGTCAACCAATGGCTGT 5821  
QY 1970 GlnValAsnTyrrAspSerCyProArgAlaIleGlnAlaGlyIleTrpTrpProArgThr 1989  
Db 5822 GAAATGAATTAATGACACTGCCCCAGACAGCATTAAGCTGGATGTGGTGGCCCTGAC 5881  
QY 1990 ArgPheGlyLeuProAlaAlaAlaProCysProLysGlySerPheGlyThrAlaValArg 2009  
Db 5882 CGCTTCGGGCTGCTGCTGCTGCTCTCTGTCCTGTCCTGTCCTGTCCTGTCCTGTC 5941  
QY 2010 HisCyAspArgLysIleArgGlyTyrrLeuProProAlaLeuPheAsnCyThrSerIleThr 2029  
Db 5942 CACTGTATGAGACAGAGGGGTGGCTCCCCCAACCTTCAATGCACTGCACTCAACCC 6001  
QY 2030 PheSerGlnLeuLysGlyPheAlaGlnArgLeuGlnArgAsnGlnSerGlyLeuAspSer 2049  
Db 6002 TTCTCAAGAACTGAAGGCTTGTGCTGATGAGGGCTGACAGCGGAATGATCAGGCTTA 6061  
QY 2050 GlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrr 2069  
Db 6062 GGGCGCTCCCAAGCACTAGCCCTGCTCTGGGCAACGGCAACCAAGCACTAGCTGCTAC 6121  
QY 2070 PheGlySerAspValLysValAlaTyrrGlnLeuAlaThrArgLeuLeuAlaHisGlnSer 2089

D 6122 TTCGAGAGGAGCTCAAGGTGGCTTACAGACTGGCCACCGCGCTGCTGGCCACGAGACC 6181  
Q 2090 ThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnLeu 2109  
D 6182 ACCCAAGCGGGCTTGTGGCTGTCTGGCCACAGAGAGTGCATTCACAGATCTGGCTG 6241  
Q 2110 ArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisPTrpGluLeuIleGlnGln 2129  
D 6242 CCGGTGGGAGCGCCCTCTGGACACAGCCACAGCGGCACTGGAGAGCTGATCCAGCAG 6301  
Q 2130 ThrGlnGlyGlyThrAlaTrpLeuLeuGlnHisPheGluAlaIleValThrAlaSerAlaLeuAla 2149  
D 6302 ACGAGAGGTGGACCGCTGGCTGGCTTCACAGACTAGAGGCTTACGCGAGTGGCTGGCC 6361  
Q 2150 GlnAsnMetArgHisPheThrLeuSerProPheThrIleValThrProAsnIleValIle 2169  
D 6362 CAGAACATGGGCACTTACCTTAAGCCCTTCAACCATGGTCAAGCCCAACATTTGCATC 6421  
Q 2170 SerValValArgLeuAspLysGlyAsnPheAlaGlyAlaLysLeuProArgTrpGluAla 2189  
D 6422 TCCGATGTCGCTTGGACAAAGGAACTTGTGGGGCCAAAGCTGCCCGCTACGAGGCC 6481  
Q 2190 LeuArgGlyGluGlnProProAspLeuGluThrThrValIleLeuProGluSerValPhe 2209  
D 6482 CTGCGTGGGAGAGAGCCCGGACCTTGAAGACAAAGTCAATTCCTTAAGTCTGTCTTC 6541  
Q 2210 ArgGluThrProProValValArgProAlaGlyProGlyGlyAlaGlnGluProGluGlu 2229  
D 6542 ACGAGAGACGCCCCCGTGGTCAAGCCCGGAGGCCCCGAGAGGCCAAGAGCCAGAGAG 6601  
Q 2230 LeuAlaArgArgGlnArgArgHisPTrpGluLeuSerGlnGlyGluAlaValAspVal 2249  
D 6602 CTGGCAGCGCGACAGAGCAGGCAACCGGAGCTGAGCCAGGGTGAAGGCTGTGGCCAGCGTC 6661  
Q 2250 IleIleTyrArgThrIleuAlaGlyLeuLeuProHisAsnThrAspProAspLysArgSer 2269  
D 6662 ATCATTTACCGACCTTGGCCGGGCTTACGCTCATACCTATGACCTTCAAGCGCACG 6721  
Q 2270 LeuArgValProLysArgProIleIleAsnThrProValValSerIleSerValHisAsp 2289  
D 6722 TTGAAGATCCCAAAAGCCGATCATCAACACCGGTGTGAGCATCAGCGTCCATGAT 6781  
Q 2290 AspGluGluLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArgLeuLeu 2309  
D 6782 GATGAGAGACTTCTGCCCCGGGCCCCGAGCAAAACCGTCAAGTGTGCGCGCTGGC 6841  
Q 2310 GluThrGluGluLysArgThrLysProIleCysValPheTrpAsnHisSerIleLeuValSer 2329  
D 6842 GAGACAGAGAGGAGGACCAAGCCCATCTGTCTTCTGGAAACCATTCATCTGTGGCAT 6901  
Q 2330 GlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAsnGluSerHisVal 2349  
D 6902 GGCAACAGGTGGCTGGTGGGCGGAGAGGCTTGAAGTGTCTTCCGCAATGAGGCCACGTC 6961  
Q 2350 SerCysGlnCysAsnHisPheThrSerPheAlaValLeuMetAspValSerArgArgGlu 2369  
D 6962 ACCTGCGAGTGCACCAATGACGAGCTTCGCTGTGCTCATGAGAGGTTTCTGGCGGGAG 7021  
Q 2370 AsnGlyGluIleLeuProLeuLysThrIleuThrThrValAlaLeuGlyValThrLeuAla 2389  
D 7022 AATGGGAGATCTGCTGACATGAGACATGACATAGGTGGCTTGAAGTGTCACTTGGCT 7081  
Q 2390 AlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHis 2409  
D 7082 GCCCTTCTGCTCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 7141  
Q 2410 GlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGly 2429  
D 7142 GGCATTCGACGATTAAGCTGACGCTGCGGCGCTGGCGCTGACGCTGCTTCTTCTTCTG 7201  
Q 2430 IleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeu 2449  
D 7202 ATCAACAGGCTGACTCTTCTTCTTCTGCTGACAGTCAATTCGCACTCTGTGCACTTCTG 7261

Q 2450 TyrLeuCysThrPheSerTrpAlaLeuLeuGluAlaLeuHisLeuTyrArgAlaLeuThr 2469  
D 7262 TACCTTGACACTTTTCTGGGCTCTGCTGAGGCTTGGACCTTACCGGGCACTCACT 7321  
Q 2470 GluValAlaArgAspValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyVal 2489  
D 7322 GAGTGGCGGAGTGCACACCGGCCCCCATGCGCTTCTACTACATCTGGGCTGGGGCGTG 7381  
Q 2490 ProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGluGlyTyrGlyLysProAsp 2509  
D 7382 CTGCGCTTCAACAGGCTAGCGCTAGCGCTGGAGCCCGAGGAGCTTACGGAAACCTGAGC 7441  
Q 2510 PheCysTrpLeuSerIleTyrAspThrLeuIleTrpSerPheAlaGlyProValAlaPhe 2529  
D 7442 TTCTGCTGGCTTCCATCTATGACAGCTCATCTGAGATTTTCTGGCCGGTGGCTTCTT 7501  
Q 2530 AlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGln 2549  
D 7502 GCCGTCTGATGAGTGTCTTCTGTACATCTGGCGGCCCCGGGCTCTCTGTGCTCCAG 7561  
Q 2550 ArgGlnGlyPheGluLysLeuGlyProValSerGlyLeuGlnProSerPheAlaValLeu 2569  
D 7562 CGGACGGGCTTGGAGAAAGATCTGTCTGGGCTGACGCTCTCTGCGCGCTCTC 7621  
Q 2570 LeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeu 2589  
D 7622 CTGCTGCTGAGGCGCAACCTGGCTGGCTGCACTGCTCTGTTCACAGGACCACTCTC 7681  
Q 2590 PheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyr 2609  
D 7682 TTCACACTACCTTTGACCTGCACTGCAATTGCAATCCAGGCCCCCTTCACTCTCTCTAT 7741  
Q 2610 ValValLeuSerLysGluValArgLysAlaLeuLysLeuAlaCysSerArgLysProSer 2629  
D 7742 GTGGGTCTTACAGAGAGGTCCGGAAGACTCAAGCTTGTGCTGAGCGGCAAGCCAGC 7801  
Q 2630 ProAspProAlaLeuThrThrLysSerThrLeuThrSerSerTyrAsnCysProSerPro 2649  
D 7802 CCGAACCTGCTGTGACCAACCAAGTCCACCTGAGCTGTCTTCAACATGCCCCAGCCCC 7861  
Q 2650 TyrAlaAspGlyAlaGluLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuHisSerThr 2669  
D 7862 TACGACATGGGCGGCTGTATCAGCCCTTACCGAATCTGCGGCTCTCTCAACAGCAC 7921  
Q 2670 SerArgSerGlyLysSerGlnProSerTyrIleProPheLeuAspGluGluSerAla 2689  
D 7922 AGTGCCTGGGAGAGATCAGCCCAAGCTACATCCCTTCTTGTGAGGAGAGATCCGCA 7981  
Q 2690 LeuAsnProGlyGlnGlyProProGlyLeuGlyLysAspProGlySerLeuPheLeuGluGly 2709  
D 7982 CTGAACCTGGCCAAAGGGCCCCCTGGCTGGGGGATCCAGGCAACCTGTTCTGGAAGT 8041  
Q 2710 GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGluAspAsp 2729  
D 8042 CAAGACCAAGCAGATGATCTTCAACGAGACTCCGACAGTGAACCTGCTTGAAGAAGCAG 8101  
Q 2730 GlnSerGlySerTyrAlaSerThrHisSerSerAspSerGluGluGluGluGluGlu 2749  
D 8102 CAGAGTGGCTCTATAGCTTCAACCACTCATCAACAGTGAAGGAGAAAGAGAGAGAG 8161  
Q 2750 GluGluGluAlaAlaPheProGlyGluGlnGlyTyrAspSerLeuLeuGlyProGlyAla 2769  
D 8162 GAAAGGAGGCGCGCTTCCCTGGAGAGACAGGCGTGGATAGCTGCTGGGCGCTGGAGCA 8221  
Q 2770 GluArgLeuProLeuLeuHisSerThrProLysAspGlyGlyProGlyLysValPro 2789  
D 8222 GAGAGACTGGCCCCCTGCAAGTACCTCCCAAGGATGGGGGCCCCAGGCGGCCGCC 8281  
Q 2790 TrpProGlyAspPheGlyThrThrAlaLysGluSerSerGlyAsnGlyValProGluGlu 2809  
D 8282 TGGCCAGAGACTTTGGGACCAACGCAAAAGATGTGTGGCAAGGCGGCCCTCGAGAGAG 8341



QY 2810 ArgLeuArgGluAsnGlyAspAlaLeuSerArgGluGlySerLeuGlyProLeuProGly 2829  
 DB 8342 CGGCTCGGGAGATGAGATGCTCTCTCCAGAGGGGCTTACGCCCCCTCCAGGC 8401  
 QY 2830 SerSerAlaGlnProHisIleGlyIleLeuGlyGlyLeuProThrIleSerGlu 2849  
 DB 8402 TCTTGGCCAGGCTCACAAGGACCTTAAAGAAAGAGTCTGCTCCACATCAGCGAG 8461  
 QY 2850 LysSerSerLeuLeuArgLeuProLeuGluGlnGlyThrGlySerSerArgGlySerSer 2869  
 DB 8462 AAGAGACGCTCTCTGGGCTCCCTCCGAGACATGACAGGCTTCTCCGGGGCTCTCC 8521  
 QY 2870 AlaSerGluGlySerArgGlyGlyProProProArgProProArgGlnSerLeuGln 2889  
 DB 8522 GCTAGTAGAGGAGCGCGGGGCGGCCCCCTCCCGCCACCGCCCGGAGAGCTCCAG 8581  
 QY 2890 GluGlnLeuAsnGlyValMetProIleAlaMetSerIleGlyAlaGlyThrValArgGlu 2909  
 DB 8582 GAGCAGCTGAAACGGGTGATGCCATCGCATGACATCAAGGCAAGCGATGATGAG 8641  
 QY 2910 AspSerSerGlySerGluPheLeuPheAsnPheLeuHis 2923  
 DB 8642 GACTCTCAGGCTCCGAATTTCTTCTTACTTCTCGAT 8693  
 RESULT 5  
 ABK15177  
 ID ABK15177 standard; DNA; 9321 BP.  
 AC ABK15177;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human REPT9 9 CDNA sequence.  
 XX  
 KW REPT9: human; antiinflammatory; cytosstatic; immunosuppressive;  
 KW antiviral; anti-HIV; antidiarrhetic; anticonvulsant; nocotropic;  
 KW neuroprotective; antiallergic; antibody; immunogen; endometriosis;  
 KW gastrointestinal disorder; gastritis; oesophageal carcinoma;  
 KW Crohn's disease; irritable bowel syndrome; ulcerative colitis;  
 KW endocrine disorder; hypochalimus disorder; Kallman's disease;  
 KW autoimmune disease; inflammatory disease; infertility; receptor;  
 KW acquired immune deficiency syndrome; AIDS; rheumatoid arthritis;  
 KW allergic; osteoarthritis; diabetes mellitus; multiple sclerosis;  
 KW systemic lupus erythematosus; cell proliferative disorder;  
 KW cancer; developmental disorder; Duchenne muscular dystrophy; gene;  
 KW Becker muscular dystrophy; neurological disorder; epilepsy;  
 KW Alzheimer's disease; Huntington's disease; reproductive disorder; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 233..9033  
 FT /tag= a  
 FT /product= "REPT9 protein"  
 FT  
 XX  
 PN MO200198354-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 21-JUN-2001; 2001MO-US19942.  
 XX  
 PR 21-JUN-2000; 2000US-214027P.  
 PR 25-AUG-2000; 2000US-228045P.  
 PR 12-DEC-2000; 2000US-255104P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Griffith JA, Kallik DA, Tribouley CM, Yue H, Nguyen DB, Tang YT,  
 PI Tai P, Policky JR, Azimzai Y, Lu DM, Graul R, Yao MG, Burford N,  
 PI Hafalia AJA, Baughn WR, Bandman O, Patterson C, Yang J, Xu Y,  
 PI Gandhi AR, Warren BA, Ding L, Sanjanwala MS, Dugan BM, Lu Y,  
 XX  
 WP1; 2002-090432/12.

DR P-PeDB; AAU74826.  
 XX  
 PT Twelve human receptors (referred to as REPT9-1 to REPT9-12), useful in  
 PT the diagnosis, treatment and prevention of gastrointestinal (e.g.  
 PT gastritis), autoimmune/inflammatory (e.g. osteoarthritis) and cell  
 PT proliferative (e.g. cancer) disorders -  
 XX  
 PS Claim 65; Page 151-154; 157pp; English.  
 XX  
 CC This invention relates to twelve human receptors CDNA sequences  
 CC referred to as REPT9-1 to REPT9-12), and the proteins encoded thereby.  
 CC The proteins of the invention may have antiinflammatory, cytosstatic,  
 CC immunosuppressive, antiviral, anti-HIV, antidiarrhetic, muscular active  
 CC general, anticonvulsant, nocotropic, neuroprotective, antiallergic  
 CC activities. The sequences of the invention may be used to produce REPT9  
 CC agonists or antagonists, and the protein sequences may be used to raise  
 CC anti-REPT9 antibodies. These molecules and the REPT9 polynucleotides and  
 CC polypeptides of the invention are useful in the diagnosis, treatment and  
 CC prevention of gastrointestinal (e.g. gastritis, oesophageal carcinoma,  
 CC Crohn's disease, irritable bowel syndrome, ulcerative colitis),  
 CC endocrine (e.g. hypochalimus disorder, Kallman's disease), autoimmune/  
 CC inflammatory (e.g. acquired immune deficiency syndrome (AIDS),  
 CC rheumatoid arthritis, allergies, osteoarthritis, diabetes mellitus,  
 CC multiple sclerosis, systemic lupus erythematosus), cell proliferative  
 CC (e.g. cancer), developmental (e.g. Duchenne and Becker muscular  
 CC dystrophy), neurological (e.g. epilepsy, Alzheimer's disease,  
 CC Huntington's disease) and reproductive (e.g. infertility, endometriosis)  
 CC disorders. Numerous other examples of each disorder are given in the  
 CC specification. The present sequence represents the human REPT9 CDNA  
 CC sequence of the invention.  
 XX  
 SQ Sequence 9321 BP; 1780 A; 3008 C; 2776 G; 1757 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 0 Length: 9321  
 Score: 15279.00 Matches: 2893  
 Percent Similarity: 98.44% Conservative: 2  
 Best Local Similarity: 98.37% Mismatches: 14  
 Query Match: 98.29% Indels: 32  
 DB: Gaps: 8  
 US-09-916-849a-3 (1-2923) x ABK15177 (1-9321)  
 QY 1 MetArgSerProAlaThrGlyValProLeuProThr---ProProProLeuLeuLeu 19  
 DB 223 ATGCGAGACCCGGGCAACCGGGGCTCCCTCCCAAGCGCGCGCCCTGCTGCTG 282  
 QY 20 LeuLeuLeuLeuLeuProProProLeuGlyAspGlnValGlyProCyAspArgSer 39  
 DB 283 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 342  
 QY 40 LeuGlySerArgGlyValArgGlySerSerGlyAlaCyAspAlaProMetGlyTyrLeuCyAspPro 59  
 DB 343 TTGGGGTCCAGGGAGACGAGGCTCTTCCGGGGGCTGCGCCCAATGGGCTGCTGCTCA 402  
 QY 60 SerSerAlaSerAsnLeuTyrTyrThrSerArgCyAspAspAlaGlyThrGluLeu 79  
 DB 403 TCCTCAGGCTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462  
 QY 80 ThrGlyHisLeuValProHisIleAspGlyLeuArgValTyrCyAspProGluSerGluAla 99  
 DB 463 ACTGGCACCTGGTACCCACACATGAGCTGAGAGGTTGGTGTCCAGAAATCCAGAGCC 522  
 QY 100 HisIleProLeuProProAlaProGluGlyCyAspProTyrSerCyAspArgLeuGluGlyTyr 119  
 DB 523 CATATTCCTCCACACACGAGCTCTGAGAGGCTGCTGAGAGCTGCTGCTGCTGCTGCTG 582  
 QY 120 GlyGlyHisLeuSerProGluGlyValLeuThrLeuProGluGlnHisIleProCyAspLeuGly 139  
 DB 583 GAGGAGCACCTTTCCTCCACAGGGCAAGCTCACCTCCCGAGAGACACCTGCTTAAAG 642  
 QY 140 AlaProArgLeuArgCyGlySerCyAspLeuAlaGlnAlaProGlyLeuArgGlyAlaGly 159



Dh 643 GCTCAAGGCTCAGATGCCAGTCTTGCAAGCTGGACAAGAGCCCGGGCTCAGGGCAGGG 702  
Qy 160 GIUATSerProgluIuSerleuGIyGIYAArgLyAspValAsnThrAlaPro 179  
Dh 703 GAAAGTCAACGAAAGTCCCTGGGTGGGCGCTCGAAGAAAGAAATGTAATACAGCCCC 762  
Qy 180 GluPheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrPro 199  
Dh 763 CAGTTCAGAGCCCCCAGGTAACCAAGCCGACAGAGTCCGGAGAACCAAGCCAGAGGCACTCC 822  
Qy 200 ValAlaSerleuArgAlaIleAspProAspGluGluAlaGlyArgleuGluTyrThr 219  
Dh 823 GTTGATCTCCGAGGCGCATGCACTGAGCCGAGCGGTGAGGCGAGTCACTGAGTACACC 882  
Qy 220 MetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGly 239  
Dh 883 ATGATGCCCTCTTGTATAGCCGCTCCAAACAGTTCCTTCTCCCTGAGCCAGTCACTGAT 942  
Qy 240 AlaValThrThrAlaGluGluIleuAspArgGluThrTyrLeuSerThrAlaPheArgVal 259  
Dh 943 GCAGTAAACCAAGCCGAGAGGTGATCGTGAACAAGAGCAACCAAGTCTTCAGGGTTC 1002  
Qy 260 ThrAlaGlnAspPheIleGlyMetProArgArgSerAlaGluAlaThrLeuThrIleLeuVal 279  
Dh 1003 ACGGCGCAGAACCAAGGATGCCGAGAGAGTGCCTGGCTACCTACCACTCATCTGGTTC 1062  
Qy 280 ThrAspThrAspAspPheAspProValPheGluGlnGlnGluTyrTyrLeuGluSerleuArg 299  
Dh 1063 ACTGACCAACCAATGACCAACCACTCTGTTCAGACGACGAGATCAAGAGAGGCTTCAGG 1122  
Qy 300 GluIleuIleuGluValGlyTyrGluValIleuThrValArgAlaThrAspGlyAspAlaPro 319  
Dh 1123 GAGAACTGAGAGTGGCTATGAGGTGCTCACTGTCAGGGCCAGGAGTGGATGCCCTT 1182  
Qy 320 ProAsnAlaAsnIleLeuTyrArgleuIleuGlySerGlySerProSerGluVal 339  
Dh 1183 CCCATGCGCAATATCTATACCGCTGCTGAGGGGCTGGGGGAGGCCCTCTGAAATCTC 1242  
Qy 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
Dh 1243 TTGAGATCGAACCTCGCTCGGGGTGATCCGAACCCCTGGCCCTGTGATCGGAGAGG 1302  
Qy 360 ValGluSerTyrGlnLeuThrValGluIleAspAspGlnGlyArgAspProGlyProArg 379  
Dh 1303 GTGGATCTCAACAGCTAGCGGTAGAGGCAAGTGAACAGGGTCGGGACCCGGGCTCTCG 1362  
Qy 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPhe 399  
Dh 1363 AGTACCAAGCCGCTGTTCTCTTCTGTGAGGATGACATGATGATGATGATGATGATGAT 1422  
Qy 400 SerGluIleAspArgValIleValIleValIleValIleValIleValIleValIleVal 419  
Dh 1423 AGTGAAGAGCCCTATGTGTCTGAGTGAAGAGATGTGATCTCCAGGGGCCCTCACTCTC 1482  
Qy 420 ArgValThrAlaSerAspArgAspGlySerAsnAlaValIleTyrSerIleMet 439  
Dh 1483 CGAGTCAAGCCCTCGATCGAGACAAAGGAGCAATGCCGTGGTGCATTAAGCATCATG 1542  
Qy 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
Dh 1543 AGTGCATATGCTCGGGGACAGTATATCTGATGCCAGACTGGAGCTCTGATATGTGTG 1602  
Qy 460 SerProleuAspArgGluThrThrTyrGluTyrThrLeuArgValAlaGlnAspGly 479  
Dh 1603 AGCCCTCTTGAATATGACACCAAGAGTACACCTTACGGGTGCGAGCACAGATGAT 1662  
Qy 480 GlyArgProProleuSerAsnValSerGlyLeuValThrValGlnValIleuAspIleAsn 499  
Dh 1663 GGCGCTCCCACTCTATATGTCTCTGCTGGTGGGACAGTACAGTCTCTGATATCAAC 1722  
Qy 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValIleuGluSerValPro 519  
Dh 1723 GACAAATGCCCATCTTATGTGATGATCAAGGCTTTCAGGCTACTGTCGTGAAGGCTGCC 1782

Qy 520 LeuGlyTyrLeuValIleuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
Dh 1783 TTAGGCTACCTGGTCTCTCATGTCCAGGCTATGCAAGCTGATGCTGGAGCAAGTCCCG 1842  
Qy 540 LeuGluTyrArgleuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
Dh 1843 CTGGAATACCGCTTGTGCGGGGTGGACATGACTTCCCTTCAACATCAATGAGGACA 1902  
Qy 560 GlyTyrIleSerValAlaIleGluIleuAspArgGluValIleAspPheTyrSerPheGly 579  
Dh 1903 GGCTGATCTCTGGGTGCTGGAACCTGACCGGAGAGAAATGATTTCTACGCTTGGG 1962  
Qy 580 ValGluAlaArgAspPheIleGlyThrProAlaLeuThrAlaSerAlaSerValThr 599  
Dh 1963 GTAAAGCTGAGACCAATGGCATCTCGAGCACTACCTCGGCGCAGTGTGAGGCTACT 2022  
Qy 600 ValIleuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
Dh 2023 GTCTTGATGTCAACGACCAATCCAACTTACCAACCAAGTACAGTGGGCTC 2082  
Qy 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAla 639  
Dh 2083 AATGAGATGCAAGCTGTGGCACCAGCGTGTGACGCTGTGACGCTGTGAGCCGTGATGCT 2142  
Qy 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
Dh 2143 CATGTGTATCACTACCAATCAAGATCCAGTGGCAATCTGAAACCCCTTCTCATACCC 2202  
Qy 660 SerGlnSerGlyGlyGlyLeuValSerleuAlaLeuProleuAspTyrTyrLeuGluArg 679  
Dh 2203 AGCCAAAGTGGTGGGTGGTATCCCTTGGCCCTGCACTGAGCTACAAACTTGAACGG 2262  
Qy 680 GlnTyrValIleuAlaValThrAlaSerAspGlyThrArgGluAspThrAlaGlnIleVal 699  
Dh 2263 CAGTATGTGTGCTGTATACCGCTCCGATGGACCTCGGAGGACCAAGGACAGATTTGTG 2342  
Qy 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
Dh 2223 GTGAATTCACCAAGCCGACCAACCCATCGTCTGTCTTCAAGACTCCCATATACGTG 2382  
Qy 720 AsnValAsnGluAspArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGlu 739  
Dh 2383 AATGTATATGAGACCGCGCGGAGGACCAAGGTGTGTGATCAAGCGCACAGATGAG 2442  
Qy 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
Dh 2443 GACACAGGTGAGATGCCGATCACTACTCATGAGGAGACAGCATCCCCAGTTCCGC 2502  
Qy 760 IleAspAlaAspThrGlyAlaValIleThrThrGlnAlaGluIleuAspTyrGluAspGlnVal 779  
Dh 2503 ATGATGACACACCGGGGCTGTACCAACCAAGCTGAGCTGAGCTACGAAAGACCAAGTG 2562  
Qy 780 SerTyrThrleuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
Dh 2563 TCTTACACCTCGGCATATATGCTCGGACAAATGGCATTCGCCAAGATCGACACACC 2622  
Qy 800 TyrleuGluIleleuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSer 819  
Dh 2623 TACCTGAGATCTGTGGGACGAGTGAACGATGAAGAATGCCCTCAATTCGCAAGACTCC 2682  
Qy 820 TyrGlnIleSerValTyrGluAspValProProPheThrSerValIleuGlnIleSerAla 839  
Dh 2683 TACCAAGGAGTGTCTATGAGATGTGCCACCTTCACTAGCTCTCGAGACTCAAGCC 2742  
Qy 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAsp 859  
Dh 2743 ACTGATGTGATTTCTGAGCTTATGCAAGGCTCTTCAACCTTCAAGAGGCGAGAT 2802  
Qy 860 GlyAspGlyAspPheIleValGluIleuSerThrSerGlyIleValIleArgThrleuAspGlu 879  
Dh 2803 GAGACGAGTACTTATATGTGATGATCAAGGCTGTGTGCAAGCTTACGAGGCTG 2862

QY 880 AsparGluAsnValAlaGlnIleuValIleuArgAlaIleuValAspIleuSerPro 899  
| | | | |  
Db 2863 GATCGAGAGAGACCTGGCCAGTATGCTTCGCGGCATATGCAATGCAAGAGGAGATGCC 2922  
QY 900 ProAlaArgThrProMetGluValIleuValIleuAspValIleuAspAsnProPro 919  
| | | | |  
Db 2923 CCGAGCCCGGACACCTATGAGAGTGAACAGTCACTGTGTGATGTGATGCAATCCCTCC 2982  
QY 920 ValPheGluGlnAspGluPheAspValPheValGluGlnAsnSerProIleuIleuAla 939  
| | | | |  
Db 2983 GTCTTTGAGCAGAGATGATGTTGATGTTGTGAGAGAGAACGCCCATGCGGCTAGCC 3042  
QY 940 ValAlaArgValIleuIleuAspProAspGluGluIleuAsnAlaGlnIleuSerIleuGln 959  
| | | | |  
Db 3043 GTGCGCCCGGCTCAAGCCATGACCCCGATGAGAGGCCCAATCCCGATTAATGATCCAG 3102  
QY 960 IleValGluGluValAsnIleProGluValPheGlnIleuAspIlePheSerGluIleuThr 979  
| | | | |  
Db 3103 ATGTGAGAGGCAACCTCCCTGAGGTCTTCACAGTGCATCTTCCTCCGCGGAGCTGACA 3162  
QY 980 AlaIleuValAspIleuAspIleuValAspArgProGluIleuValIleuValIleuAlaIleu 999  
| | | | |  
Db 3163 GCCCTGTAGACTTAACTAGAGAGACCGGCTGAGTACCTCTGTGATCAAGGCCACG 3222  
QY 1000 SerAlaProIleuValSerArgAlaIleuValIleuValArgIleuAspArgAsnAspAsn 1019  
| | | | |  
Db 3223 TCAAGCTCTCTGTGAGCCGGGCTACAGTCACTCCGCTCTTGAACCGCATGACAAAC 3282  
QY 1020 ProProValIleuGluValAsnPheGluIleuPheAsnAsnIleuIleuAsnAsnSer 1039  
| | | | |  
Db 3283 CCAACAGATGCTGGGCAACTTGAATCTTTCAACAACATATGCAACAATCCCTCAAC 3342  
QY 1040 SerPheProGluValAlaIleuValProAlaIleuAspProAspIleuSerAspSer 1059  
| | | | |  
Db 3343 AGCTTCCTGGGGGCTGCAATGGCCAGATCACTGCCATGACCTGATATCTCAAGTAA 3402  
QY 1060 LeuThrTySerPheGluArgGluValAsnGluIleuSerIleuValIleuAsnAlaSerThr 1079  
| | | | |  
Db 3403 CTGACTTACAGCTTGGAGCGGGAAATGAATCACTCACTGGTCTCTGATGCTCCACG 3462  
QY 1080 GlyGluIleuValIleuSerArgAlaIleuAspAsnAspArgProIleuIleuAlaIleuMetSer 1099  
| | | | |  
Db 3463 GGTGAGCTGAGCTAAAGCCGCGCATGAGCAACAACCGGCTCTGAGAGGCATCAAGAGC 3522  
QY 1100 ValIleuValSerAspGluValIleuSerValIleuGlnCysAlaIleuArgValIleuIleu 1119  
| | | | |  
Db 3523 GTGCTGGTGTCAAGCCGCTACACAGGCTGACCGCCCAAGTCCGCTGCTGACATC 3582  
QY 1120 IleThrAspGluMetIleuThrIleuSerIleuIleuArgIleuValAspMetSerProGlu 1139  
| | | | |  
Db 3583 ATCAACCATGAGATGCTCAACCAAGCATCACTGCGCTGAGAGGACATGTCAACCCGAG 3642  
QY 1140 ArgPheIleuSerProIleuGluIleuPheIleuAlaValAlaIleuIleuAlaIleu 1159  
| | | | |  
Db 3643 GCGTTCCTGACCACTGCTAGGCTCTTCACTCAAGCGGAGGCGGCAAGCTGAGCAG 3702  
QY 1160 ProProAspIleuValIleuPheAsnValGlnIleuAspThrAspAlaProGluIleuIleu 1179  
| | | | |  
Db 3703 CCAACCGGACCAAGGTGTCTTCAAGTCAAGGAGGACCAACGCGCCCGGAGGCGCAC 3762  
QY 1180 IleIleuAsnValSerIleuSerValGluGlnProProGluIleuIleuProPhe 1199  
| | | | |  
Db 3763 ATCTCAACAGTGAAGCTGTGTGTGTGGCAGACCGGCAAGGCGCGGCGGCGCCCTTC 3822  
QY 1200 LeuProSerGluAspIleuGlnIleuArgIleuIleuIleuAsnArgSerIleuIleuThrAlaIleu 1219  
| | | | |  
Db 3823 CTGCGCTCTTGAAGACCTGAGAGCGGCTTAACTCAACCGCAAGCTGCTGAGAGGCATC 3882  
QY 1220 SerAlaGlnIleuValIleuProPheAspAspAsnIleuCysIleuArgIleuProCysGluAsn 1239  
| | | | |  
Db 3883 TCGGCAACAGGCGGCTGCGCTTGAAGCAACATCTGCGGAGGAGCCCTCGAGAGAC 3942  
QY 1240 TyMetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleuAlaSerSer 1259  
| | | | |

Db 3943 TACATGCGGCTGTGCTCGGTGCTGCGCTTCACTCTCGCGGCTTATGAGCTCTCC 4002  
| | | | |  
QY 1260 SerValIleuPheArgProIleuIleuProValGluIleuValArgCysArgCysProProGlu 1279  
| | | | |  
Db 4003 TCCGTGCTCTTCCGCGCATCAACCCGCTGAGAGGCTGCGCTGCGCGCGCGCGC 4062  
QY 1280 PheThrGluAspIleuTySerGluIleuValIleuAspIleuValIleuAspArgProCysGluPro 1299  
| | | | |  
Db 4063 TTCAAGGCTACTGCGAGACCGAGGTGAGTCACTCTCTACTCCGCGGCTCTGAGCC 4122  
QY 1300 HisGluArgCysArgSerArgGluGluIleuIleuTyThrCysIleuCysArgAspIleuTyThr 1319  
| | | | |  
Db 4123 CACGGGCGGCTGCGGACCGGAGCGGCGGCTACACTGCTCTGTGATGATGATGACAG 4182  
QY 1320 GlnGluIleuIleuValSerAlaArgSerGluValArgCysIleuProGluIleuValCysIleuAsn 1339  
| | | | |  
Db 4183 GGTGAGCACTGTGAGGTGAGTCTGCTGACAGCCCTTCACCCCGGTGTCTGCAAGAT 4242  
QY 1340 GlyIleuThrCysValAsnIleuValGluIleuPheIleuValAspCysAspCysProSerGluValAsp 1359  
| | | | |  
Db 4243 GGGGCACTGTGTCAACTGCTGTGGCGGCTTCAAGTGCATGCGCATTCGATGAGAC 4302  
QY 1360 PheGluIleuProTyCysGlnValIleuThrArgSerPheProAlaIleuSerPheIleuThr 1379  
| | | | |  
Db 4303 TTCAGAGAGCCCTACTGCAAGGTGACACGCGACCTTCCCGCCCATCTTCAAC 4362  
QY 1380 PheArgGluIleuArgGlnIleuArgPheIleuPheIleuAlaIleuSerPheAlaIleuIleu 1399  
| | | | |  
Db 4363 TTTCCGCGGCTGCGGCAAGGTTCCACTTCACTCCGCTCTCTGCTTGGCCCAAGAGAG 4422  
QY 1400 ArgAspGluIleuIleuIleuTyAsnGluValArgPheAsnGluValAspPheValAlaIleu 1419  
| | | | |  
Db 4423 CCGCAAGGCTGTCTGTGTATCAATGGCGCTTCAATGAGAGCATGCTTGTGCGCTC 4482  
QY 1420 GluValIleuGlnIleuValGlnIleuThrPheSerAlaGluIleuSerThrThrVal 1439  
| | | | |  
Db 4483 GAGGTGATCCAGAGCAGGTCACTCACTTCTGCAAGGAGATCAACCAACAGGAG 4542  
QY 1440 SerProPheValProGluIleuValSerAspGluIleuIleuPheIleuValGlnIleuValTy 1459  
| | | | |  
Db 4543 TCCCAATTCGTGCGGAGAGATCAATGATGCGCATGAGTCACTGCAATAC 4602  
QY 1460 TyAsnIleuProIleuIleuGluIleuIleuIleuProGlnIleuProSerGluIleuIleuVal 1479  
| | | | |  
Db 4603 TACATTAAGCACTGTGTGCTGACAGGGCTCCACAGGGCCATCAAGAGCAAGAGT 4662  
QY 1480 AlaValIleuIleuValAspGluCysAspThrGluValAlaIleuArgPheGluSerValIleu 1499  
| | | | |  
Db 4663 GCTGTGTGACCGTGTGATGCTGTGACACAGAGTGTGCTGCGCTGCGATCTGCTG 4722  
QY 1500 GlyAsnTySerCysAlaAlaGlnIleuIleuIleuIleuIleuIleuIleuIleuIleu 1519  
| | | | |  
Db 4723 GGCACACTCTCTGTGCTGCGCCAGGCAACCGAGGTGCGAGAGATCTTGAATCTG 4782  
QY 1520 ThrGluProIleuIleuGluIleuValProAspIleuProIleuSerPheProValArgMet 1539  
| | | | |  
Db 4783 ACGGGGCCCTGTACTAGAGCGGGGTGCTGACCTGCGCCGAGAGCTTCCAGTCCGAATG 4842  
QY 1540 ArgGlnPheValGluCysMetArgAsnIleuIleuValAspSerArgHisIleuAspMetAla 1559  
| | | | |  
Db 4843 CCGCAGTGTGTGTGCTGACATGGAACCTGCAAGGTGACAGCGGCACTAATACATGCT 4902  
QY 1560 AspPheIleuAlaAsnAsnGluIleuValProGluIleuProAlaIleuValAsnValCysAsp 1579  
| | | | |  
Db 4903 GACTTCATTGCGCAACATGAGCACTGCTGTGCTGCTGCTGCGCAAGAGATGATGATC 4962  
QY 1580 SerAsnThrCysHisAsnGluIleuIleuValAsnGlnIleuIleuIleuIleuIleuIleu 1599  
| | | | |  
Db 4963 ACCCACTTGTGCAATGAGGCGCACTTGTGAGACAGGTGAGAGCGCTTCAAGCTGCAAG 5022  
QY 1600 CysProIleuGluPheGluIleuValAsnSerCysAlaGlnIleuAlaAsnProGluIleuIleu 1619  
| | | | |

Db	5023	GGCCCCCTTGGGGGCGAAGAGCTGGCCGACGAAATGCCAATCCACAGACCTTC	5082
Qy	1620	LeuGIySerSerLeuValAlAlATrPhIseGIyLeuSerLeuProIIseGIyProTrPTyI	1639
Db	5083	CTGGGACACAGCTGGTGGCCCTGGCATGGCCCTCTGCTGGCCCATCTCCCAACCCCTGGTAC	5142
Qy	1640	LeuSerLeuMetPheArgTrThrArgGlnIaAapPGIyValIleuLeuGlnAlIeThrArg	1659
Db	5143	CTAGACCTCAATGTTCCGACCGCGCCAGCCAGCGGTCTCTGCTGACGAGCCATCACCAAG	5202
Qy	1660	GIyArgSerThrIleThrLeuGlnLeuArgGlnIyHsValMetLeuSerValGlnGIy	1679
Db	5203	GGGCGCCACACACCATCACCTTACAGCTACAGAGGGGCCACGTGAGCTGAGGGTGAAGGGC	5262
Qy	1680	ThrGIyLeuGlnAlaIaSerSerLeuArgLeuGIyUPProGIyAaAaAaArgGIyAaPTTr	1699
Db	5263	ACAGGGCTTCAGAGCTCTCTCTCTCGTGTGAGCCAGGGCCGAGCCAAAGAGGTGACTGG	5322
Qy	1700	HisHisIaGIyLeuAlaLeuGlnIyAlaSerGIyGIyProGIyHsAlaIleLeuSerPhe	1719
Db	5323	CACCATGCACAGCTGGGACCTGGAGGCCAGCGGGGGGCGCTGGCCATGGCATTCCTCTCTTC	5382
Qy	1720	AapTrYGlyGlnGlnAlaArgAlaGlnGIyAaMetLeuGIyProAlaGleuHisGlyLeuHisLeu	1739
Db	5383	GATTATGGGACAGACAGAGAGAGAGGGCAACTGGGGCCCCGGGCTGCATGGTCTCACCTG	5442
Qy	1740	SerAaIIeThrValGIyGIyIleProGIyProAlaGIyGIyValAlaArgGIyPheArg	1759
Db	5443	AGCAACATMAAGTGGGGCGGAAATCTGGCCAGCCGGCGGGTGGGCCCGCTTTCGG	5502
Qy	1760	GIyCybLeuGlnGIyValArgValSerAaPTThrProGIyGIyValAaSerLeuAaPTPro	1779
Db	5503	GGCTGTTTGGCAGGGGTGGTCGGGTGGCATTCGCCACAGGGGGGTGTTAAACGCTGATCC	5562
Qy	1780	SerHisGIyGlnSerIleAaSerValGlnGlnGIyCybSerLeuProAaPTProCybAaPTSer	1799
Db	5563	AGCCATGGGAGAGCATATACGTGAGCAAGGCTGTAGCCCTGACCTTGACCTTGACCTCA	5622
Qy	1800	AaPTProCybProAlaAaSerTrCybSerAaAaPTTPAaPTSerTrYSerCybSerCyb	1819
Db	5623	AAACCGTGTCTCGCTAAACAGCTAATGACGAACAGCATGGGACAGCATATTCCTCGACGCTG	5682
Qy	1820	AaPTProGIyTrTrTrGIyAaAaPTCybThrAaPTValCybAaPTLeuAaPTProCybGlnHis	1839
Db	5683	GATCCAGGTTACTATGTGTGACAACTGTACTATGTGTGTGACCTGTGAACCCGTGTGACAC	5742
Qy	1840	GlnSerValCybThrArgIyAaPTProSerAlaProHisGIyTrThrCybGlnCybProPro	1859
Db	5743	CAGCTGTGTGTATCCCGGAACCCAGTGGCCCCCATGGCTATACCTCGAGTGTCCCCCA	5802
Qy	1860	AaPTTrLeuGIyProTrTrCybGlnThrArgIleAaPTProCybProATGTrITPTTr	1879
Db	5803	AATTAACCTTGGGCCATCTGTGAAACCGAGTTTACCAAGCCCTTGTCCCGGTGGCTGGTG	5862
Qy	1880	GlyHisPTProTrCybGlnProCybAaPTCybAaPTValSerIyGlyPheAaPTProAaPTCyb	1899
Db	5863	GGACATCCCAATGTGGCCACAGACCTGTATGTCAAGAAAGGCTTTGACCCAGACTGC	5922
Qy	1900	AaPTTrSerSerGIyGlnCybHisCybIyGlnAaHisGTrYAaPTProGIySerPro	1919
Db	5923	AACAAGACAAACGGCGAGTGCCTACGTGAAGAAACCACTACCGGGCCCCCAGGAGGCC	5982
Qy	1920	ThrTrCybLeuLeuCybAaPTCybTrTrProThrGIySerLeuSerAaPTValCybAaPTProGln	1939
Db	5983	ACCTGCTCTGTGTGATGTGTACCCCAAGAGCTCTTGTCTCAAGACTGTGTACCCCTGAG	6042
Qy	1940	AaPTGIyGlnCybProCybIyAaPTProGIyValIleGIyArgGlnCybAaPTCybAaPTAa	1959
Db	6043	GATGGCCAGTGTCCATGTGAACCAAGGTGTATCTGGGGGTCACTGTGACCCCTGTGACAC	6102
Qy	1960	ProPheAlaGlnValThTrAaPTCybGln-----ValAaPTTrAaPTSer	1975
Db	6103	CCTTTTCGTAGGTGACCAACAAATGGCTGTGTGAAGGGCCCTGTGTTGTGATGAC-----	6156

[illegible]

QY	2326	lleuValSerGIYThrGIYThrSerIAlaArgGIYCySGIValAlaIleuPheArgAsn	2335
Dp	7225	ATCTGTGATGATGGACAAGATGGCTGGTCGGCACAAGGCTGTGAAGTCGTCTCCGCAAT	7284
QY	2346	GIuSerHISValSerCYSGIInCyASnMI;SWetThSerPheAlaValIleuMetAspVal	2365
Dp	7285	GAGAGCCACGTCAAGCTGGCAAGTGAACCAATGACGAGCTTCGGCTGCTCATGAGCGTT	7344
QY	2366	SerArganArgIU-----AenGIValIleu	2374
Dp	7345	TCTGGCGCGAGGTCGGGCCCAAGAGGGAGCTGACAGCGGTGGAAGGGAGAAATCTGTG	7404
QY	2375	ProIeuIYThrIleuThrTYrValAlaIeuGIYValThrIleuAlaAlaIleuIleuThr	2394
Dp	7405	CCAATGAAGACATGACATACGTGGCTCTTAAGTGTCACTTGCGCTCCCTCTCTGCTACAC	7464
QY	2395	PhePhePheIleuThrIleuIleuArgIIleuIArgSerAenGIInIAGIYIleArgArgAsn	2414
Dp	7465	TTCTCTCTTCCATCACTCTGTGGTATCTCGGCGCTCCAAACAAGCGATCCGACGTATAC	7524
QY	2415	IeuThrAlaAlaIeuGIYIleuAlaGIInIeuValPheIleuIeuGIYIleAenGIAlaIAsp	2434
Dp	7525	CTGACAGCTGCGCTGGGCTGGCTGAGCTGAGTCTTCTCTCGGGAATACACAGGCTGAC	7584
QY	2435	IeuProPheAlaCYeThrValIIeAlaIleIleuIeuMI;SphIeUryIleuCYeThrPhe	2454
Dp	7585	CTCCCTTTTGGCTGCAAGTACATTCGCANCCGCTGACCTTCGTATCCTCGACCTTT	7644
QY	2455	SerTrrpAlaIleuIeuGIuAlaIeuMI;SleuTYrArgAlaIleuThrGIuValIArgSphVal	2474
Dp	7645	TCTGGGCTGTGCTGGAGAGCCTTGACCTGACCTGAGCACTCAAGATGCGCGCATGTC	7704
QY	2475	AenThrGIYProMetArgPheTYrTYrMetIeuGIYTPGIYValProAlaPheIIethr	2494
Dp	7705	AACACCGGCCCCATGCGCTTGTACATCAATGCTGGGCTGGGCGTGGCTTCATCAACA	7764
QY	2495	GIYIleuAlaValGIYIleuAspProGIuGIYTYrGIYAspProAspPheCYeTrrpIleuSer	2514
Dp	7765	GGGCTAGCGGTGGGCTGTGAGCCGCCAGAGGACTACAGGAAACCTGTACCTTCTGGCTCC	7824
QY	2515	IleTYrAspTrrpIleuIleTrrpSerPheIAGIYProValAlaPheAlaValSerMetSer	2534
Dp	7825	ATCTATGACACCGCTCATCTGGAAGTTTGTGTGGCCGGGTGGCTTTGCCGTCTCATAGAGT	7884
QY	2535	ValPheIleuTYrIIleuAlaAlaArgAlaSerCYAlaIAGIArgGIInIYpHeGIu	2554
Dp	7885	GTCCTTCTGTACATCTGGCGGCGCGGCGCTCTGTGCTGCCCAAGCGCAAGGCTTTGAG	7944
QY	2555	LYsIYArgIYProValSerGIYIleuInProSerPheAlaValIleuIleuIleuSerIa	2574
Dp	7945	AAGAAAGGTCGTCTCGGCGCTGTGAGCCCTCTTCGGCGTCTCTCTGTGTAGGCCCC	8004
QY	2575	ThrTrrpIleuIleuAlaIleuIleuSerValIAsnSerAspTrrpIleuIleuPheMI;eTYrIleuPhe	2594
Dp	8005	ACGTGGCTGTGGGACGTGCTCTCTGTCAACAAGCAGCAACCTCTCTTCACTACACTCTTT	8064
QY	2595	AlaThrCYeAsnCYeIIeGIInIYProPheIIePheIleuSerTYrValIalleuSerIYS	2614
Dp	8065	GCTACTCTCAATGTGATCCAGGGGCCCTTCACATCTTCTCTTAATGTGTGTTCAGAG	8124
QY	2615	GIuValIArgIYValaIleuIYAlaIeuIYAlaCYeSerArgIYProSerProAspProAlaIeu	2634
Dp	8125	GAGGTCCGGAAGACACTCAAGCTTGTGTGACGCCGACAGCCGCTGACCTGTCTGTG	8184
QY	2635	ThrThrIYSerThrIleuThrSerSerTYrAenCYProSerProTYrAlaAspGIYArg	2654
Dp	8185	ACCACCAAGTCCACCTGACCTGTGCTTCAACACTGCCACAGCCCTTACGCAAGATGGGCGG	8244
QY	2655	IeuTYrGIInProTYrGIYAspSerIAGIYSerIleuMI;AserThSerArgSerGIYYS	2674
Dp	8245	CTGTATCAAGCCCTTACAGAGACTGGCGGCTTCTGTGACAGACCAAGTGTCTGGGCAAG	8304
QY	2675	SerGIInProSerTYrIIeProPheIleuIArgGIuIleuSerAlaIeuAsnProGIYGIIn	2694

[illegible]

XX antifungal; vulnereary; antiulcer; ss.  
 XX Homo sapiens.  
 XX WO200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 XX  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HSE-) HXSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-457740/49.  
 DR P-PDB; ABB11404.  
 XX  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 XX  
 XX Claim 1; Page 496-498; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoietic regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness.  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.  
 CC  
 XX  
 SQ Sequence 9401 BP; 1839 A; 2961 C; 2644 G; 1957 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 9401  
 Score: 13378.50 Matches: 2542  
 Percent Similarity: 99.14% Conservative: 2  
 Best Local Similarity: 99.06% Mismatches: 12  
 Query Match: 86.06% Indels: 10  
 DB: 22 Gaps: 3

US-09-916-849a-3 (1-2923) x ABA08648 (1-9401)  
 QY 368 GLUAlaserAPGInGIYARGAAPProGIYProArg-SerThrThrAlaAlaValPheLe 387  
 DB 3 GAGGCAAGTGAACAGAGGCTGGGACCCGGGTCTCGGAGTACCAACAGCGCTGTTTCT 62  
 QY uSerValGIuAspAspAspAspAspAlaProGlnPheSerGIuYAspArgTYValAlGI 407  
 DB TTCTGTGGAGAGTACACATATATATGCCCCCACTTATGTGAAGAGCCATATGTGTCCA 122  
 QY nValArgGIuAspValThrProGlyAlaProValIleuArgValThrAlaSerAspArgAs 427  
 DB GGTGAGGAGAGATGTGACTCCAGGGGCCCGAGTCCGAGTCAAGCTCGATCCAGAGA 182  
 QY pIyAGLISerAspAlaValAlaHisTYrSerTleMerSerGIYAspAlaArgGIYGINPh 447  
 DB CAAAGGAGCAATCCGCTGGTGACTATAGCATATAGTGGCAATGCTCGGGAGACAGTT 242  
 QY eTYrIleuAspAlaGlnThrGIYAlaIleuAspValIaSerProIleuAspTYrGIuThrTh 467  
 DB TTATCTGGATGCCAGACTGGAGCTGTGANTGTGTAGACCTTGTACTATAGAGACAC 302  
 QY IYleGIuTYrThrIleuArgValArgAlaGlnAspGIYGIYArgProIleuSerAspAs 487  
 DB CAAAGAGTACACCTTACGGGTGGAGAGCAGAGATGTGGCCGCTCCCACTCTTAATGT 362  
 QY lSerGIYleuValThrAlaGlnIleuAspAlaAspAspAlaProIlePheValIse 507  
 DB CTCTGGCTTGTGACATGACAGTCTGGAGTATGAAGACAAATCCCACTTGTGTGAG 422  
 QY rThrProPheGlnAlaThrValIleuGIuSerValProIleuGIYTYrIleuValIleuHisVa 527  
 DB CACCCCTTCCAGGCTACGTGCTGGAGAGGTCCCTTGGGCTTACCTGGTTCCATGT 482  
 QY lGlnAlaIleAspAlaAspAlaGlyAspAspAlaArgIleuGIuTYrArgIleuAlaGIYVa 547  
 DB CCAAGCTATGACACCTATGTGTGTGCAATGCCCGCTGGAAACCGCTTGTGGGT 542  
 QY lGIYHisAspPheProPheThrIleAspAspGIYTYrIleSerValAlaAlaGI 567  
 DB GAGACATGACTTCCCTTCCATCAATCAATGAGCAGAGCTGTGCTGTGGCTGCTGA 602  
 QY uIleuAspArgGIuGIuValAspPheTYrSerPheGIYValGIuAlaArgAspHisGIYTh 587  
 DB ACTGACCGGAGAGAAATTATTTCTACAGCTTGGGTGAAAGCTCGAGACCATGGCAC 662  
 QY rProAlaIleuThrAlaSerAlaSerValIleuValIleuAspValAspAspAs 607  
 DB TCCAGCACTGACTGCTCGGCAAGTGTACAGCTGACCTGTGATGTCAACGACAA 722  
 QY nProThrPheThrGlnProGIuTYrThrValArgIleuAspGIuAspAlaAlaValGIYTh 627  
 DB TCCAACTTTTACCACCAACAGTACAGTGCAGCTGCTCAATGAGATGACAGCTGGGAC 782  
 QY rSerValValThrValSerAlaValAspArgAspAlaHisSerValIleThrTYrGINI 647  
 DB CAGGCTGTGACGGGTGCACTGTGACCGTGAATGCTCATATGTCACTTACCCAGAT 842  
 QY eThSerGIYAsnThrArgAspArgPheSerIleThSerGIuSerGIYGIYIleuVal 667  
 DB CACCAAGTGGCAATCTGAAACCGCTTCTCATCAACCAAGCAAGGTGGTGGGCTGGT 902  
 QY lSerIleuAlaIleuProIleuAspTYrIYIleuGIuArgGIYTYrValIleuAlaValThrAl 687  
 DB ATCCCTTGCCCTGCACTGACCTACAACTTGAACCGGCACTATGTGTGGCTGTACCGC 962  
 QY aSerAspGIYThrArgGlnAspThrAlaGlnIleValIaAspValThrAspAlaAsnTh 707  
 DB CTCCGATGAGCACTCGGAGAGACAGGACAGATGTGTGAATGTCAACCGCAACAC 1022  
 QY rHisArgProValPheGlnSerSerHisTYrThrValAspValAspGIuAspArgProAl 727

Db 1023 CCATGCTCTGCTTTACAGCTCCCACTATACAGTAATGTTATATGAGGACCGGCGGC 1082  
Qy 727 aglyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAsnAlaArgI 747  
Db 1083 AGGCACACAGGATGCTGTATACAGCGCACGATGAGGACACAGGTGAGAAATCCCGCAT 1142  
Qy 747 ethrTyrrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspThrGlyAlaVa 767  
Db 1143 CACCTCTTCATGTAGAGACAGCATCCCGCAGTTCGCCATCATGATGACAGACACGGGCGCTGT 1202  
Qy 767 IThrThrGlnAlaGluLeuAspTyrrGluAspGlnValSerTyrrThrLeuAlaIleThrAl 787  
Db 1203 CACCAACCAAGCTGAGCTGACATACAGAACACAGGTCTTACACCTGCGCATTACTGC 1262  
Qy 787 aArgAspAsnGlyIleProGlnIleSerAspThrThrTyrrLeuGlnIleLeuValAsnAs 807  
Db 1263 TCGGAGCAATAGGCAATCCCAAGAAAGTCCGACACCACTACCTGAGAAATCTGTGTAGACA 1322  
Qy 807 pValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrrGlnGlySerValTyrrGluAs 827  
Db 1323 CGTGAATGACAAATGCCCCCTCAAGTTCCTGCGAGACTCTTACCAAGGCGCATGTCTATAGGA 1382  
Qy 827 pValProProPheThrSerValLeuGlnIleSerAlaThrAspArgAspSerGlyLeuAs 847  
Db 1383 TGTGCAACCTTCTACTAGGCTCTGCAATCTCAGCCACTGATCGATTCGTGACTTAA 1442  
Qy 847 nglyArgValIlePheTyrrThrPheGlnGlyGlyAspAspGlyAspGlyAspPheIleValG 867  
Db 1443 TGGCAGAGGCTTCTTCAACCTTCCAGAGGCGACGATGAGACGGTGAACCTTTATGTTGA 1502  
Qy 867 userThrSerGlyIleValIleArgThrIleuArgArgLeuAspArgGluAsnValAlaGlnTy 887  
Db 1503 GTCCACAGTACGGATGTGTGCGAACGCTACGAGGCTGATCGAGAACGGTGGCCACAGTA 1562  
Qy 887 rValIleuArgAlaTyrrAlaValAspIleGlyMetProProAlaArgThrProMetGluVa 907  
Db 1563 TGTCTTGCGGGCATATGACGTGACAGAGGAGATGCCCGACGCCGACACTTATGAAAT 1622  
Qy 907 IThrValIThrValIleuAspValIAsnAspAsnProProValPheGlnGluAspGluPheAs 927  
Db 1623 GACAGTCACGTGTGTGATGTGAAATGACATCCCTGTCTTGAACAGATGAGTTGA 1682  
Qy 927 pValPheValIleGluIleuAspSerProIleGlyLeuAlaValAlaArgValThrAlaThrAs 947  
Db 1683 TGTGTTTGGAGAGAACACAGCCCCCATGTGGCTAGCGCGGCTGACAGCCACTGA 1742  
Qy 947 pProAspGluGlyThrAsnAlaGlnIleMetTyrrGlnIleValIleGluGlyAsnIleProG 967  
Db 1743 CCCCAGTGAAGGACCAATGCCAGATTAATGCCAGATTTGTGAGGGGCAACATCCCTGA 1802  
Qy 967 uValPheGlnLeuAspIlePheSerGlyGlyIleuThrAlaLeuValAspLeuAspTyrrG 987  
Db 1803 GGTCTTCCAGCTGAGCATCTTCTCCGGGAGCTGACAGCCCTGTGTGACTTGAATACGA 1862  
Qy 987 uAspArgProGlnTyrrValIleuValIleGlnAlaThrSerAlaProLeuValSerArgAl 1007  
Db 1863 GACCGGCGCTGAGTACGTCTGTGATCCAGGCAAGTCAAGCTCTCTGTGTAGACCGGCGC 1922  
Qy 1007 aThrValIleValIleuArgLeuAspArgAsnAspAsnProProValIleuGlyAsnPheG 1027  
Db 1923 TACAGTCCACGTCCGCTCTTGAACCGCAATGACCAACCAACGAGTCTGGGCAACTTTGA 1982  
Qy 1027 uIleLeuPheAsnAsnTyrrValIThrAsnArgSerSerSerPheProGlyGlyAlaIleG 1047  
Db 1983 GATCCCTTTCACAACTATGTACCAATGCTCAACACCACTTCCCTGGGGGTGCCATTGG 2042  
Qy 1047 yArgValProAlaIleAspProAspIleSerAspSerLeuThrTyrrSerPheGluArgG 1067  
Db 2043 CCGAGTACCTGCGCAATGACCTGATATCTCAGATACCTGACTTACAGCTTTGAGGGGG 2102  
Qy 1067 yAsnGluLeuSerLeuValIleuLeuAsnAlaSerThrGlyGluLeuIleLeuSerArgAl 1087  
Db 2103 AATGAATCTACGCTGTGTCTGTCAATGCTTCCAGGGGTGAGCTGAAGTCAAGCGGCGC 2162

Qy 1087 AleuAspAsnAsnArgProLeuGlnAlaIleMetSerValIleuValSerAspGlyValAla 1107  
Db 2163 ACTGACAAACAACCGGCTCTGAGAGGCATATAGAGCTGTGTGTCAACCGGCTTACA 2222  
Qy 1107 sSerValIThrAlaGlnCyAlaAlaLeuArgValIThrIleIleThrAspGluMetLeuThrHi 1127  
Db 2223 CAGCGTACCGGCGGCTGCGGCTGTGTGATCCATTCATCAACCGATVAGATGTCTCACCA 2282  
Qy 1127 sSerIleThrLeuArgLeuGluAspMetSerProGluArgPheLeuSerProLeuLeuG 1147  
Db 2283 CAGCATACGCTGCGCTGTGAGACATATCAACCGAGGCTTCTCTGTACCACTGCTAGG 2342  
Qy 1147 yLeuPheIleGlnAlaValAlaAlaThrLeuAlaThrProProAspAlaValValPh 1167  
Db 2343 CCTTTCATCCAGGGGGGTGGCCGCGCACGCTGGCCACCGCACCGAGTGTGTCTT 2402  
Qy 1167 eAsnValGlnArgAspThrAspAlaProGlyGlyHiAlaIleLeuAsnValSerLeuSerVa 1187  
Db 2403 CAACGTACAGCGGGAACACCGACGCCCGGGGGCCACATCTTCAACGTGACCTGTGCGT 2462  
Qy 1187 IGlIleProProGlyProGlyGlyGlyProProPheLeuProSerGluAspLeuGlnG 1207  
Db 2463 GGGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2522  
Qy 1207 uArgLeuTyrrLeuAsnArgSerLeuLeuThrAlaIleSerAlaGlnArgValIleuProPh 1227  
Db 2523 GCGCTTATACCTCAACCGCACCTGTGACGGCCATCTCGGACACGCGGTGTGCTCTT 2582  
Qy 1227 eAspAspAsnIleCyAlaLeuArgGluProCyArgGluAsnTyrrMetArgCyValSerValLe 1247  
Db 2583 CACACACAAATCTGTCTGTGAGAGCCCTGTGAGAACTCAATGCGCTGTGTGCGTCT 2642  
Qy 1247 uArgPheAspSerSerAlaProPheIleAlaSerSerSerValIleuPheArgProIleHi 1267  
Db 2643 GGGCTTGAATCTCTCCGGGCGCTTATATGCGCTTCTCTCGTGTGCGCTTCCGGCCATCA 2702  
Qy 1267 sProValIleGlyLeuArgCyAlaArgCyAspProGlyPheThrGlyAspTyrrCyArgIleTh 1287  
Db 2703 CCCCCTCGAGAGGCTGCGCTCCGCTGCGGCGGCTTCAACGGGTGACTTCTGTGAGAAC 2762  
Qy 1287 rGluValAspLeuCyAspTyrrSerArgProCyArgIleProHiAglyArgCyAspSerArgG 1307  
Db 2763 CAGGTGAGACCTGTGCTACTGTGCGGCTGTGTGAGCCCGACCGGCGCTCCGACCGCGGA 2822  
Qy 1307 uGlyGlyTyrrThrCyAspLeuCyAspAspGlyTyrrThrGlyGluHiAspGluValSerAl 1327  
Db 2823 GGGGCGGCTACACGTGCTCTGTGTGTGATGCTACCGGCTGAGCACTGTGAGTGTAGTGC 2882  
Qy 1327 aArgSerGlyArgCyThrProGlyValCyAlaAsnGlyGlyTyrrCyValAsnLeuLe 1347  
Db 2883 TCGCTCAGGCGGTTGCAACCGGCGGTGTGTGCAAGATGGGGGCACTGTGTCAACTTCT 2942  
Qy 1347 uValIleGlyPheIleAspCyAspAspCyProSerGlyAspPheGlyIleuTyrrCyArgGlnVa 1367  
Db 2943 GGTGGGCGGTTCAGATGCGATGTGCCATCTGTGAGACTTCAAGAAAGCTTACTGCAAGT 3002  
Qy 1367 IThrThrArgSerPheProAlaIleSerPheIleThrPheArgGlyLeuArgGlnArgPh 1387  
Db 3003 GACCAACGCGGAGCTTCCCGGCGCACTCTTCATCACTTTCGCGGCGTCCGCGCACGCTT 3062  
Qy 1387 eHiPheThrIleuAlaLeuSerPheAlaThrTyrrGluArgAspGlyLeuLeuLeuTyrrAs 1407  
Db 3063 CCACTTCAACCTGGCGCTCTGTGTTTCCACAAAGAGGCGACAGGTTGTGTGTATAA 3122  
Qy 1407 nglyArgPheAsnGlnuysHiAspPheValAlaLeuGlnValIleGlnGlnGlnValAlG 1427  
Db 3123 TGGGCTTTCATAGAGACATGACTTTGTGCGCTCGAGGTATCCAGAGCAAGTCTCA 3182  
Qy 1427 nLeuThrPheSerAlaGlyGluSerThrThrValSerProPheValProGlyGlyVa 1447  
Db 3183 GCTCACTTCTCTGAGGAGGATCAACCAACAGGTGTGCCCATTCGTGCGCGGAGGAGT 3242



1447 | 1serAapGlyIntPrh1SerValGlnLeuLeuYrTyraaLeuYProLeuLeuGlyG1 1467  
3243 | CAGTATGCGCAGTGGCATACGGTGCAGCTGAATCTCAATTAAGCCACTGTGGGTCA 3302  
1467 | nThrGlyLeuProGlnGlyProSerGlnGlnValAlaValThrValAspGlyCy 1487  
3303 | GACAGGGCTCCACAGAGGCCCATCAGAGCAGAGAGGCTGTGTGACCGTGTATGGCTG 3362  
1487 | aAspThrGlyValAlaLeuArgPheGlySerValLeuGlyAanTySerCysAlaG 1507  
3363 | TGAACAGAGAGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 3422  
1507 | IngTyThrGlnGlyGlySerLeuYLeuSerLeuAapLeuThrGlyProLeuLeuGlyG 1527  
3423 | AAGGGCACCCAGGGGTGGCACAAGAACTCTGGATCTGACGGGGCCCTGCTACTAGGGG 3482  
1527 | ValProAapLeuProGlySerPheProValArgMetArgGlnPheValGlyCysMetA 1547  
3483 | GAGTGTCTAAGCTGGCCGAGAGCTTCCAGTCCGAATGCGGAGTTCGTGGCTGCATGC 3542  
1547 | rGAsnLeuGlnValAAspSerArgHisLeaAspMetAlaAspPheLeaAlaAsnGlyT 1567  
3543 | GGAACCTGAGGTGGACAGCCGGCACAATGACATGCTACTTCAATTCACCAATGGCA 3602  
1567 | hValProGlyCysProAlaValValYAsnValCysAAspSerAanThrCysHisAsnGlyG 1587  
3603 | CGGTGCTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3662  
1587 | YrTyThrCysValAanGlnTTPAAspAlaPheSerCysGlyCysProLeuGlyPheGlyGly 1607  
3663 | GACTGTGCTGAGAACAGTGGAGAGCGCTTCAAGCTGCAGAGTCCCTGGCTTGGGGGCA 3722  
1607 | ySerCysAlaGlnLeuMetAlaAspProGlnHisPheLeuGlySerSerLeuValAlaT 1627  
3723 | AAGCTGCGCCCAAGAGAAATGGCCCAATCCACAGACTTCTGGGGCAGACCTGTGGCTT 3782  
3783 | GGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3842  
1647 | rGAlaAlaAspGlyValLeuLeuGlnAlaHisLeuThrArgValArgSerThrHisLeuLeuG 1667  
3843 | GCGAGGCCCAAGGAGTGTCTGCTGCAAGCCATCCACAGAGGGGGCGCAGACCAATCCTTAC 3902  
1667 | hLeuAAspGlyGlnHisValMetLeuSerValGlnGlyThrGlyLeuGlnAlaSerSerL 1687  
3903 | AACTAGAGAGGGCCACGTGATCTGAGCTGAGGGGCAAGGGCTTCAAGGCTCTCTTC 3962  
1687 | euAArgLeuGlnProGlyArgAlaAsnAapGlyAAspTrpHisHisAlaGlnLeuAlaLeuG 1707  
3963 | TCCGTCTGAGGCGAGCGCGGCAATGAGCGTGAAGCCATGCAAGCTGGCACTGG 4022  
1707 | ValAspSerGlyGlyProGlyHisAlaHisLeuSerPheAAspTyArgGlnGlnAlaGAlaG 1727  
4023 | GAGCCATCGGGGGGCTGCGCAGTCCATTCGTCTTCAATTAAGGCGCAGAGAGCG 4082  
1727 | IngTyAanLeuGlyProArgLeuHisGlyLeuHisLeuSerAanHisLeuThrValGlyGlyT 1747  
4083 | AAGGGCACTGGGGCCCGCGCTGCAAGTCTGCACTGAGCAACATTAAGTGGCGGAA 4142  
1747 | LeProGlyProAlaGlyValAlaAaArgGlyPheArgGlyCysLeuGlnGlyValArgV 1767  
4143 | TACCTGGGCGACCGCGGTGTGGCCGTGGCTTTCGGGGCTGTGGCGAGGTGTGGCGG 4202  
1767 | aLeSerAAspThrProGlnGlyValAAspSerLeuAAspProSerHisGlyGlnSerHisAanV 1787  
4203 | TGAAGCATTCGCCAAGAGGGGGTTTACAGCTGAGATTCACAGCCATGGGAGACATCAAG 4262  
1787 | aGlnGlnGlnGlyCysSerLeuProAAspProCysAAspSerAanProCysProAlaAsnSerT 1807  
4263 | TCGAGCAAGAGCGTGTAGCTGCTGCACTTGTGTGACTCAAAACCGTGTCTGTCAAGCT 4322  
1807 | YrCysAAspAAspTrpAAspSerTySerCysSerCysAAspProGlyTyTyArgValAspA 1827

4323 | ATTCAGAGCAAGACTGGAGAGCTATTCCTGACAGTGTATCCAGGTTACTATGTGTACA 4382  
1827 | enCysThrAanValCysAAspLeuAAspProGlyHisGlnSerValCysThrArgGlySP 1847  
4383 | ACTGTACTTAATGTGTGTACTGTAAACCGTGTGTAGCACAGTGTGTGTGTGTGTGTGTGT 4442  
1847 | rSerAlaProHisGlyTyTyThrCysGlyCysProProAAspTyLeuGlyProTyCysG 1867  
4443 | CCAAGTGGCCCAAGGCTTACTGACAGTGTCCCAATTAACCTTGGGCAATCTGTG 4502  
1867 | hThrArgHisAAspGlnProCysProArgGlyTyTrpGlyHisAProThrCysGlyProC 1887  
4503 | AGACAGAGATTGACACAGCTTGTCCCGGTGTGGTGGAGATCCACATGTGGCCCAT 4562  
1887 | yAAsnCysAAspValSerTyGlyPheAAspProAAspCysAanTyThrSerGlyGlnCysH 1907  
4563 | GCAACTGTATGTACAGCAAGAGCTTGTACCCAGACTGTACACAGCAGCGGAGTGC 4622  
1907 | LeCysLeuGlnAAsnHisTyTyArgProProGlySerProThrCysLeuLeuCysAAspCysT 1927  
4623 | ACTGAGAGAGAACCACTACCGGCCCCAGAGCAGCCCACTGCTTGTGTGTACTCT 4682  
1927 | YrProThrGlySerLeuSerArgValCysAAspProGlnAAspGlyGlnCysProCysGlySP 1947  
4683 | ACCCCAGAGGCTCTGTGTCCAGAGTGTGTGACCTGTAGAGATGGCCAGTGTCCATGCAAGC 4742  
1947 | rGlyValHisGlyValArgGlnCysAAspArgCysAAspAAspProPheAlaGlnValThrThrA 1967  
4743 | CAGGTGTATGGGGGTGAGTGTACCGCTGTGTACMACCTTGTGTGTGTGTGTGTGTGTGT 4802  
1967 | snGlyCysGlyValAAsnTyTrpAAspSerCysProArgAlaHisGlnAlaGlyTLeuTrpTP 1987  
4803 | ATGCTGTGAAGTGAATTAATTAACAGCTGCCACAGAGCAATTAAGGCTGGATCTGTGTGC 4862  
1987 | rAArgThrArgPheGlyLeuProAlaAlaAlaProCysProGlyGlySerPheGlyThrA 2007  
4863 | CCGGTACCGGCTGGGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4922  
2007 | ValAlaArgHisCysAAspGlnHisAAspGlyTyTrpLeuProProAAspLeuPheAAspCysThrS 2027  
4923 | CTGTGCGCACTGTATGTAGCAAGAGGGGTGTCTCCCCCAACCTTCAACTGTCAAGT 4982  
2027 | erHisLeuPheSerGlyLeuLeuGlyPheAlaGlnAArgLeuGlnAArgAAspGlnSerGlyL 2047  
4983 | CCATACCTTCTCAAGAACTGAAGGGCTTGTGTGAGCGCTTCAAGGGAATGATCAGGCC 5042  
2047 | euAAspSerGlyArgSerGlnGlnLeuAlaLeuLeuLeuAAspAlaThrGlnHisThrA 2067  
5043 | TTAGCTCAGGCGGCTCCACAGCAGTACCTGTCTTCCGCGCAAGCGCCACGACGACACAG 5102  
2067 | LagTyTyrPheGlySerAAspValHisValAlaTyArgGlnLeuAlaThrArgLeuLeuAlaH 2087  
5103 | CTGGCTACTTGGGAGCAGCTCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5162  
2087 | LeGlnSerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAAspValHisPheThrGlnA 2107  
5163 | ACGAGAGACCCACAGGGGCTTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5222  
2107 | enLeuLeuAAspValGlySerAlaLeuLeuAAspThrAlaAsnTyArgHisTyTrpGlnLeuI 2127  
5223 | ATCTGTGCGGGGT 5282  
2127 | LeGlnGlnThrGlnGlyGlyTyThrAlaTrpLeuLeuGlnHisTyArgGlnAlaTyArgAlaSerA 2147  
5283 | TCCAGAGACAGAGAGGT 5342  
2147 | LeuAlaGlnAAspMetArgHisTyTyThrTyLeuSerProPheThrHisValThrProAanI 2167  
5343 | CCTGTGCGCCAGAACATGTGGGACACCTTACCTTAAGCCCTTCAACATGTGTGTGTGTGTGT 5402  
2167 | LeValHisSerValValArgLeuAAspTyGlyAanPheAlaGlyValAlaTyLeuProAArgT 2187



Db 5403 TTGTCATCTCCGAGTGGCGCTTGACAAAGGAATTGCTGGGGCCAAAGTCCCGCT 5462  
Qy 2187 YTCGUAAlaLeuAArgIyGluGlnProProAspLeuGluThrThrValIleLeuProGluS 2207  
Db 5463 ACCAGGCGCTGCTGGGGAGACAGCCCCGGAACCTTAGACAAAGATCTTCTGCTGAGT 5522  
Qy 2207 eValIpheArgIuThrProProValValArgProAlaGlyProGlyGluAlaGlnGluP 2227  
Db 5523 CTGCTTTCAGAGAGAGAGCCCCCGGTGGTCAAGCCCGAGGCCCCGAGAGGCCCAAGAG 5582  
Qy 2227 rGluGluLeuAlaAArgArgGlnAArgHisProGluLeuSerGlnGlyGluAlaValA 2247  
Db 5583 CAGAGAGCTGGACAGCGGACAGACAGAGCCGAGCTGAGCCAGGGTGAAGGCTGTGG 5642  
Qy 2247 IAsSerValIleIleIyYrArgThrIleuAlaGlyLeuLeuProHisAsnYrAspProAspI 2267  
Db 5643 CCAAGCTCATCATCTACCGAACCTTGGCCGGGCTTACGCTCATTAATAGACCTTGACA 5702  
Qy 2267 yArArgSerLeuAArgValProLyAspProIleIleAsnThrProValValSerIleSerY 2287  
Db 5703 AGCGCAGCTTGAGAGATCCCAAGCCGATCAACACACCCGTGTGAGCATCAGCG 5762  
Qy 2287 aHisAspAspGluGluLeuLeuProArgAlaLeuAspLySProValThrValGlnPhea 2307  
Db 5763 TCCATGATGATGAGAGAGCTTCTGCCCCGGGCGCTGACAAACCGTCAAGGTCAAGTTCC 5822  
Qy 2307 rGluLeuGluGluThrGluGluAArgThrLyAspProIleCySValPheTrpAsnHisSerIle 2327  
Db 5823 GCTGCTGAGAGACAGAGAGCCGACCAAGCCCATCTGTCTTCTGGAACCATTCATCC 5882  
Qy 2327 euValSerGlyThrGlyGlyTyrSerAlaArgGlyCySValValPheArgAsnGluS 2347  
Db 5883 TGGTCAAGTGGCACAGGTGGCTGGCGGACAGAGCTGTGAAGTGGTTCGCGAAATGAGA 5942  
Qy 2347 eHisIleValSerCySValGlnCySAsnHisSerThrSerPheAlaValIleuMetAspValSer 2367  
Db 5943 GCGACGTCAGCTGCAAGTGCACACACAGAGCTTGGCTGTGATGAGAGGTTCTC 6002  
Qy 2367 rGArArgIuAenGlyGluIleLeuProLeuYrSerThrLeuThrYrValAlaLeuGlyValA 2387  
Db 6003 GGGGGAGATGGGGAGATCTGCTGCACTGAAGACATGACATGCTGGCTCTAGGTGTCA 6062  
Qy 2387 hIleuAlaAlaLeuLeuLeuThrPhePheLeuLeuThrLeuLeuArgIleLeuArgSerA 2407  
Db 6063 CCTTGGCTGCTCTGCTGCTCACTTCTTCTGCTCACTTCTTGGCGATCTGGCTGCCA 6122  
Qy 2407 snGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPhe 2427  
Db 6123 ACCAACCGGCAATCCAGCTGACCTGACAGCTGCGGCTGGGCTGAGCTGAGTGTCTCC 6182  
Qy 2427 euLeuGlyIleAsnGlnAlaAspLeuProPheAlaCySThrValIleAlaIleLeuLeu 2447  
Db 6183 TCCGAGGATCAACCAAGCTGACCTCTTTGGCTGACACAGTCAATGGCATCTGTGTC 6242  
Qy 2447 IAsPheLeuYrLeuCySThrPheSerTrpAlaLeuLeuGluAlaLeuHisLeuYrArgA 2467  
Db 6243 ACTTCCGTGACCTCTGACCTTTTCTGGGCTGTGAGAGGCTTGCACTGTACCGGG 6302  
Qy 2467 IAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheYrYrMetLeuGly 2487  
Db 6303 CACTCATGAGGTGCGGAGATCAACACCGGCCCAATGCGCTTCTTACATGATGCTGGCT 6362  
Qy 2487 rPgIyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGluGlyYrGly 2507  
Db 6363 GGGGCGTGGCTCTTATCAACAGGCTAGCCGGGCGCTGAGACCCGAGAGGCTACCGGA 6422  
Qy 2507 snProAspPheCySThrLeuSerIleYrAspThrLeuIleTrpSerPheAlaGlyProY 2527  
Db 6423 ACCGTGACTTCTGTGCTCTCCATCTATGACAGGCTCATCTGAGTTTGTGGCCCGG 6482  
Qy 2527 aAlaIlePheAlaValSerMetSerValPheLeuYrIleLeuAlaAlaArgAlaSerCyS 2547  
Db 6483 TGGCTTTTGGCGTCTCATGATGATCTTCTCTGATCATCTGGGCGCCCGGCGCTCTGTG 6542

Qy 2547 IAlaGlnAArgGlnGlyPheGlyLySValGlyProValSerGlyLeuGlnProSerPhea 2567  
Db 6543 CTGCCACAGCGAGGCGCTTGAAGAGAGATCCGTCTCGGCGCTGACGCCCTCTTCC 6602  
Qy 2567 IAlaIleuLeuLeuSerAlaThrTrpLeuAlaLeuLeuSerValAsnSerAspT 2587  
Db 6603 CCGTCTCTGCTGTGAGCGCCACGTGAGCTGTGCACTGCTCTCTGTCAACAGCA 6662  
Qy 2587 hIleuLeuPheHisThrLeuPheAlaThrCySAsnCySValGlnGlyProPheIlePhe 2607  
Db 6663 CCTCTCTCTCACTACCTCTTGTACTGATGATGATCCAGAGGCCCTTCACTTCC 6722  
Qy 2607 euSerYrValValLeuSerLySValArgValAlaLeuLySLeuAlaCySAspArgL 2627  
Db 6723 TCTCTGATGTGTGCTTGAAGAGAGTCCGGAAGACATCAAGCTTGTGCTGACCGCA 6782  
Qy 2627 yAsProSerProAspProAlaLeuThrThrLySAspThrLeuThrSerSerYrAsnCyS 2647  
Db 6783 ACCCAGCGCTGACCTCTGTCGACACACAGTCCACCTGACCTGTCTTCAACTGTC 6842  
Qy 2647 rGArProYrAlaAspGlyArgLeuYrGlnProYrGlyAspSerAlaGlySerLeu 2666  
Db 6843 CAGGCCCTTACGCAATGGCGGCTGTACAGCCCTTAACGAGACTCGGCGGCTCTGTG 6902  
Qy 2667 HisSerThrSerArgSerGlyLySAspGlnProSerYrIleProPheLeuLeuArgIu 2686  
Db 6903 CACAGACCAAGTCCGTCCGGGACAGAGTCAAGCCCACTCATCTCTTGTGTGAGGAG 6962  
Qy 2687 GluSerAlaLeuAspProGlyGlnGlyProProGlyLeuGlyAspProGlySer 2704  
Db 6963 AGTCCGCACTGACCTTGGGCGCAAGGCGCCCTTGGGCGGATTCGAGAGAGG 7022  
Qy 2705 LeuPheLeuGluGlyGlnAspGlnHisAspProAspThrAspSerAspSe 2722  
Db 7023 TGTTCCTGGAGAGGTTCAAAGACAGAGCATTAATCTTAACAGAGATTTCAGACAG 7082  
Qy 2722 rAspLeuSerLeuGluAspAspGlnSerGlySerYrAlaSerThrHisSerSerAspSe 2742  
Db 7083 TGAACGTCTTAAAGAGACAGACAGATGCTCTTAATGCTTACCCACTCATCAGACAG 7142  
Qy 2742 rGluGluGluGluGluGluGluGluGluGluAlaPheProGlyGluGlnGlyTrpAs 2762  
Db 7143 TAAAG 7202  
Qy 2762 pSerLeuLeuGlyProGlyValaGluArgLeuProLeuHisSerThrProLyAspGlyG 2782  
Db 7203 TAGCTGTGGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7262  
Qy 2782 yProGlyProGlyLySAlaProTrpProGlyAspPheGlyThrThrAlaLySLeuSerSe 2802  
Db 7263 CCAAGGCGCTGGCAAGGCGCCCTGGCCAGAGACATTTGGAGACACAGCAAAAGAGGTG 7322  
Qy 2802 rGlyAsnGlyAlaProGluGluArgGluGluAsnGlyAspAlaLeuSerArgGluG 2822  
Db 7323 TGGCAAGGAGGCGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7382  
Qy 2822 ySerLeuGlyProLeuProGlySerSerAlaGlnProHisLySValIleLeuLySLe 2842  
Db 7383 GTTCCAGGCGCCCTTCCAGGCTTCTTGGCCAGGCTCAAAAGGATCTTAAGAGAA 7442  
Qy 2842 sCySLeuProThrIleSerGlyLySAspSerLeuLeuArgLeuProLeuGluGlnCyS 2862  
Db 7443 GTGTGCGCCACATCAG 7502  
Qy 2862 rGlySerSerArgGlySerSerAlaSerGlyGlySerArgGlyGlyProProProArgPr 2882  
Db 7503 AGGATTTTCCCGGGCTCTTCCGCTAGTGAAGGAGAGAGAGAGAGAGAGAGAGAG 7562  
Qy 2882 oProProArgGlnSerLeuGlnGluLeuAsnGlyValMetProIleAlaMetSerI 2902  
Db 7563 ACCGCCCGGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 7622

QY 2902 elysalaglythralaspGluaspSerSerGlySerGlyluPheleuPhePheAsnPhle 2922  
 DB 7623 CAGGCGAGCACGCTGATGATGAGGATCGTCAGGCTCCGATTTCTCTTTAACTTCTT 7682  
 QY 2922 whis 2923  
 DB 7683 GCAT 7686

RESULT 7  
 AAC76401  
 ID AAC76401 standard; cDNA; 9121 BP.  
 AC AAC76401;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORF1956 polynucleotide sequence SEQ ID NO:3911.  
 XX  
 KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;  
 KW vulnery; antipapillary; antipapilloma; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antipapillary; immunosuppressant; cardiac;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antichyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinketsu RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR P-PsDB; AAB42192.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 3061-3067; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORF open reading frames 1 to 3161. The ORF  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipapillary; antipapilloma; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antipapillary; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antichyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 9121 BP; 1797 A; 2892 C; 2529 G; 1901 T; 2 other;

Alignment Scores:  
 Pred. No.: 0 Length: 9121  
 Score: 12839.50 Matches: 2413  
 Percent Similarity: 98.74% Conservative: 23  
 Best Local Similarity: 97.81% Mismatches: 28  
 Query Match: 82.60% Indels: 3  
 DB: 21 Gaps: 2

US-09-916-849a-3 (1-2923) x AAC76401 (1-9121)

QY 457 AspValIserProLeuAspPyrGluThrThrLysGlyThrThrLeuArgValArgAla 476  
 DB 4 GATGGTGAGCCCTTGTACTATGAGCAGCAGAGTAACCTTACGGGTGGACCA 63  
 QY 477 GlnAspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeu 496  
 DB 64 CAGAGTGTGGCCGTCCTCCCTCTTATGTCCTGTGGCTTGGTACAGTACAGTCTG 123  
 QY 497 AspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlu 516  
 DB 124 GATATCAACGCAATATGCCCATCTTCCTGACAGCCCTTTCAGGGCTACTGCTGAG 183  
 QY 517 SerValProLeuGlyThrLeuValLeuHisValGlnAlaIleAspAlaGlyAsp 536  
 DB 184 AGCTGCCCTTATAGCTACTGTTCTTCATCTCAGGCTATGACGCTGATGCTGTGAC 243  
 QY 537 AsnAlaArgLeuGlyThrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsn 556  
 DB 244 AATGCCGCTGAGATATACCGCTTGTGGGTGGAGCATGACTTCCCTTACCATCAAC 303  
 QY 557 AsnGlyThrGlyThrPylIserValAlaAlaGluLeuAspArgGluGluValAspPheThr 576  
 DB 304 AATGACACAGGCTGATCTCTGTGGCTGTGACATGACCGGAGGAAATTTCTTAC 363  
 QY 577 SerPheGlyValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerVal 596  
 DB 364 AGCTTGGGGTGAAGCTCAGACCATGCGACTCAGACCTCATGCTCGGCCAGTGTG 423  
 QY 597 SerValThrValLeuAspValAsnAspAsnAsnProThrPheThrGlnProGluThr 616  
 DB 424 AGCGTACCTGCTGATGTCACGACGACCAATCCACTTACCCCAACGAGATGACCA 483  
 QY 617 ValArgLeuAsnGluAspAlaAlaValAlaGlyThrSerValValThrValSerAlaValAsp 636  
 DB 484 GTGGGGCTCAATGAGGATGCGAGCTGTGGGACCGCGTGGAGCGGTGACGTGTGAC 543  
 QY 637 ArgAspAlaHisSerValIleThrThrGlnIleThrSerGlyAsnThrArgAsnArgPhe 656  
 DB 544 CGTATGCTCATATGTCATCATCACTTACCATGCTCAGTGCATTCGAAACCGCTTC 603  
 QY 657 SerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspPyrIys 676  
 DB 604 TCCATTCACGACCAATATGGGTGGGTGGCTGTATCCCTTGGCCCTGACATGACATCAAA 663  
 QY 677 LeuGluArgGlnThrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAla 696  
 DB 664 CTGAGCGGCGATGATGATGCTGTATCCGCTCCAGTGCCTCGGACGACGACGCA 723  
 QY 697 GlnIleValValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHis 716  
 DB 724 CAGATTGGTGATGATGACGACGACGACCAACCATGCTGTCTTTCAGAGCTCCAC 783  
 QY 717 TyrThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAla 736

```

Db 784 TATACAGTGAATGTTATGAGAACCGGCCGACAGCACACGCTGGTGTGATCAGCCGCC 843
QY 737 ThrAspGluAspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIlePro 756
Db 844 ACCGATGAGGACACAGGTGAGAAATGCCCGATCACCTTCACTTCAATGAGAGACACATCCCC 903
QY 757 GluPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGlu 776
Db 904 CAGTTCCGATCGATCGAGACACGGGGGCTGTCAACACCGAGCTAGAGCTACGACGAA 963
QY 777 AspGlnValSerTyrThrIleuAlaIleThrAlaArgAspAsnGlyIleProGlnIlySer 796
Db 964 GACCAAGTGTCTTACACCTCGGCCATTACTGTCTGGGACAAATGGCAATCCCCAGAAAGTCC 1023
QY 797 AspThrThrTyrLeuGluIleLeuValAspAspValAspAspAsnAlaProGlnPheLeu 816
Db 1024 GACACCACTTACCTGGAGATCTGTGTGAACGAGTGAATGAATGATGCCCTCAGTTCTCG 1083
QY 817 ArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerValLeuGln 836
Db 1084 CGAGACTCTTACACAGGGCAGTGTCTTATGAGGATGTGCACCTTCACTAGCGTCTGCAAG 1143
QY 837 IleSerAlaThrAspArgAspSerGlyLeuAsnGlyAlaArgValPheTyrThrPheGlnGly 856
Db 1144 ATCTCAGCCACTGATCGTGAATCTGAGCTTAATGCGAGGGCTCTTCAACCTTCCAAAGGA 1203
QY 857 GlyAspAspGlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeu 876
Db 1204 GGGGAGATGTGAACCGGTGACTTATGTGTGAGTCCAGTCAAGCAATCGTCCAAAGCTA 1263
QY 877 ArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIlyS 896
Db 1264 CGAGAGCTGTGATCGAGAGAACTGTGGCCAGTATGTCTTGTGGGCAATGTGAGTGGCAAG 1323
QY 897 GlyMetProProAlaArgThrPrometGluValThrValThrValLeuAspValAsnAsp 916
Db 1324 GGGATGCCCCCAGCCCGCACCTATGGAAGTGAACATCACTGTGTGAGTGAATGGAATGAC 1383
QY 917 AsnProProValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIle 936
Db 1384 AATCCCCCTTCTTGTGACAGATGATGATGTGTGTGTGTGTGAGAGAGAAACAGCCCAT 1443
QY 937 GlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIle 956
Db 1444 GGGCTAGACCGTGGCCGGGTCAACAGCCATGACCCCGATGAAAGGACCAATGCCAGATT 1503
QY 957 MetTyrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGly 976
Db 1504 ATGTACCAAGATTGTGGAGGCAATCTCGAGGTCTTCCAGCTGGACATCTTCTCGGG 1563
QY 977 GluLeuThrAlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIle 996
Db 1564 GAGCTGACAGCCCTGTGAGATCTTGAAGGACCGGCTCTAGATCGTCTGTGATC 1623
QY 997 GlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArg 1016
Db 1624 CAGGCCACGTCAAGCTCTGTGTGAGCGGGCTACAGTCCACGTCCCGCTCTTGAACCGC 1683
QY 1017 AsnAspAsnProProValIleGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsn 1036
Db 1684 AATGACCAACCAACAGTGTGGGCACTTGAAGTCTTTTCAACAACATATGTCACCAAT 1743
QY 1037 ArgSerSerPheProGlyGlyAlaIleGlyAlaArgValProAlaHisAspProAspIle 1056
Db 1744 CGCTCAAGACGCTTCCCTGGGGGTGCCATTGGCCGAGATCCTGTCCATGACCCGATATC 1803
QY 1057 SerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsn 1076
Db 1804 TCAATATGTCTGACTTACAGCTTGTGAGCGGGAAATGAACCTCAGCTGTGCTCAT 1863
QY 1077 AlaSerThrGlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAla 1096

```

```

Db 1864 GCCTCCACGGGTGACGTGAAGCTTAAGCCGCGCATGGAACAAACAGCCGCTTGGAGGCG 1923
QY 1097 IleMetSerValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArg 1116
Db 1924 CTATGAGGATGTCTGTGTGTATGACATCCACAGCGTCAAGGCTTCTTGCACCTTGCT 1983
QY 1117 ValThrIleIleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMet 1136
Db 1984 GTCAACATCATCAACGAGACATGCTGACCAACAGCATCATCTGCTCGCTGAGAAACATG 2043
QY 1137 SerProGluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThr 1156
Db 2044 TCCCAAGAAAGTTCCTGTCCCGCTGGCCCTTCTGTGAGAGGGGTGAGCCGCGTG 2103
QY 1157 LeuAlaThrProProAspHisValValPheAsnValGlnArgAspThrAspAlaPro 1176
Db 2104 CTGTCCACCAACCAAGACAGCTCTTGTCTTCAACGTCCAGAACGACACCGACGTC--- 2160
QY 1177 GlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGly 1196
Db 2161 AGCTCCACATCTGAAACGTGACTTCTCGCGCGCTGCTGCGCGCGCTCCGCGGC--- 2217
QY 1197 ProProPheLeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeu 1216
Db 2218 ---CAGTTCTTCCGTCCGAGGACCTGCAAGAGAGATCTTGAATCGAACGCTGTG 2274
QY 1217 ThrAlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluPro 1236
Db 2275 ACCACATCTTCAACGACAGGCGCTGCTCCCTTCAACGACAAATCTGCTGTGGCGAGGCC 2334
QY 1237 CysGluAsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIle 1256
Db 2335 TCGAGAACTACATGAAGAGCGTGTCCGTTCGCAATTCGACGCTCCCGCCCTTCCTC 2394
QY 1257 AlaSerSerSerValLeuPheAspProIleHisProValGlyIleLeuArgCysArgCys 1276
Db 2395 AGCTCCACAGGTGTCTTCCGCGCATCCACCCCATACAGGCTGCGCTGCGCTCC 2454
QY 1277 ProProGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgPro 1296
Db 2455 CGGCCCGGCTTACCGGGGACATCTGCGAGACGAGATGACCTGTGCTACCTCGCGCGG 2514
QY 1297 CysGlyProHisGlyAlaArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAsp 1316
Db 2515 TGGGGGCCAAAGCGCGCGCTCGCAGCGCGAGGGCGGTACACCTGCTGTGTGAT 2574
QY 1317 GlyTyrThrGlyGluHisCysGluValSerAlaArgSerGlyAlaArgCysThrProGlyVal 1336
Db 2575 GGCTACACGGGTGAGCACTTGAAGTGAAGTCTCGCTCAGGCCGTGGCACCCCGGCTGTC 2634
QY 1337 CysIlyAsnGlyGlyThrCysValAsnLeuLeuValGlyIlyPheIlyCysAspCysPro 1356
Db 2635 TCGAAGAAATGGGGGACCTGTGTCAACCTGTGTGGGGGTTTCAAGTGCATGCCCCA 2694
QY 1357 SerGlyAspPheGlnIlyAspProTyrCysGlnValThrThrArgSerPheProAlaHisSer 1376
Db 2695 TCTGAGACTTGAAGAGCCCTTACCTGCAAGTGAACACCGAGCTTCCCGGCCACTCC 2754
QY 1377 PheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAla 1396
Db 2755 TTCAATACCTTTCGGCGGCTGCGGCGAGCTTTCACCTTCACTGACCTGCGCTCTGTGGC 2814
QY 1397 ThrIlyGluArgAspGlyLeuLeuLeuTyrAsnGlyAlaArgPheAsnGluIlyHisAspPhe 1416
Db 2815 ACAAAAGAACCGGACGCGGTGTGTGTCATATGGCGGCTTCAATGAAGAAGCATGACTTT 2874
QY 1417 ValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThr 1436
Db 2875 GTGGCCCTGAGGTGATCCAGAGAGAGTCAAGCTCACTTCTTCAAGGAGGTCAACC 2934
QY 1437 ThrThrValSerProPheValProGlyGlyValSerAspGlyGlnTyrHisThrValGln 1456
Db 2935 ACCAGGTGTCTCCCATTTGTCGCGCGGAGAGTGAAGTGAAGTGCAGTACGAGTGCAG 2994

```

QY 1457 LeuLYSTYRAsnLYsPProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGln 1476  
 Db 2995 CTGAATATCTACAAATAGCCACTGTGGGCTCAGACAGGGGCTCCACAGGGCCATCAGAG 3054  
 QY 1477 GlnLYsValAlaValValThrValAspGlyCyAspThrGlyValAlaLeuArgPheGly 1496  
 Db 3055 CAGAAAGTGGCTGTGGTGAACCGTGAATGGCTGTGACACAGAGAGTGGCTTGGCTTCGGA 3114  
 QY 1497 SerValLeuGlyAsnLYsSerCyAspAlaGlnGlyThrGlnGlyGlySerLYsLYsSer 1516  
 Db 3115 TCTGTCTTGGGCACTACTCTCTGTCTGTCCAGGGCAACCAGGGTGGCAGCAAGAGTCT 3174  
 QY 1517 LeuAspLeuThrGlyProLeuLeuGlyGlyValProAspLeuProGlyLeuSerPhePro 1536  
 Db 3175 CTGGATCTGACCGGGGCCCCCTGCTCACTAGGCGGGGTGGCTGACTGGCCCGAGAGCTTCCA 3234  
 QY 1537 ValArgMetArgGlnPheValGlyCyMetArgAsnLeuGlnValAspSerArgHisIle 1556  
 Db 3235 GTCCCAATGGCGGCACTTCTGTGGCTGTGACGCGAACTGTGAGTGTGACAGCGGCACTA 3294  
 QY 1557 AspMetAlaAspPheIleAlaAsnAsnGlyThrValProGlyCyAspProAlaLYsLYsAsn 1576  
 Db 3295 GACATGGCTGACTTACTTCCCAACATGGCACCGTGTCTGGCTGTCCCTGCGCAAGAAAGAC 3354  
 QY 1577 ValCyAspSerAsnThrCyAspHisAsnGlyGlyThrCyAspValAsnGlnTrpAspAlaPhe 1596  
 Db 3355 GTGTGTGACAGCAACACTTGCACATGGGGGCACTTGGGTGTGACAGTGGGAGCGGTTC 3414  
 QY 1597 SerCyAspGlyCyAspProLeuGlyPheGlyGlyLYsSerCyAspAlaGlnLeuMetAlaAsnPro 1616  
 Db 3415 AGCTCGAGTGTCCCCCTGGGCTTGTGGGCAAGACTGGCCGCGCAAGAAATGGCCAAATCCA 3474  
 QY 1617 GlnHisPheLeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGln 1636  
 Db 3475 CAGCACTTCTTGGGAGAGAGCTGTGGCTGTGGCATGGCTCTCTGCTCCATCTCCCA 3534  
 QY 1637 ProTrpLYsLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAla 1656  
 Db 3535 CCTGTGACTCTCAGGCTCACTATGTTCCGACAGCGCCGAGGGCGGAGGTGCTGTGACAGGCC 3594  
 QY 1657 IleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSer 1676  
 Db 3595 ATCACAGAGGGGCGGAGCACATCACTCAAGCTACAGAGAGGGCCAGCTGATGTGTAGC 3654  
 QY 1677 ValGlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAsp 1696  
 Db 3655 GTGGAGGCGCACAGGGCTTCAAGGCTCTCTCTCTCTGTGTGAGCCAGGCCGGGCCAAATGAC 3714  
 QY 1697 GlnAspTrpHisHisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 Db 3715 GGTCGCTGGCACATGACAGCTGGGCACTGGGAGCGCAGGGGGGCTGTGGCAATCCATT 3774  
 QY 1717 LeuSerPheAspTrpGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHisGly 1736  
 Db 3775 CTGTCTCTTGANTATAGGGGACAGAGACAGAGGGCAACTGGGGCCCCGGCTGTGATGTT 3834  
 QY 1737 LeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaArg 1756  
 Db 3835 CTGCACCTGAGCAACATACAGTGGGCGGAATCTTGGGCGAGCCGGGGGTGTGGCCCT 3894  
 QY 1757 GlyPheArgGlyCyLeuGlnGlyValAlaArgValSerAspThrProGlnGlyValAsnSer 1776  
 Db 3895 GGCTTTCGGGGCTTGTTCAGGGGTGTGGGTGAGACGATACGCCAGAGGGGGTTAAACAGC 3954  
 QY 1777 LeuAspProSerHisGlyGlyLeuSerIleAsnValGlnGlnGlyCySerLeuProAspPro 1796  
 Db 3955 CTGAATCCCAAGCATGGGAGAGATCAACGTGGAGCAAGGCTGTGCTGTGACCTT 4014  
 QY 1797 CyAspSerAsnProCyAspProAlaAsnSerLYsSerAspAspTrpAspSerLYsSer 1816  
 Db 4015 TGTGACTCAAAACCGGTGTCTGTCTAAAGCTATTTCAGCAACGACTGGGACAGGCTATTCC 4074

QY 1817 CysSerCyAspProGlyLYsTrpGlyAspAsnCyThrAsnValCyAspLeuAsnPro 1836  
 Db 4075 TGCAGCTGTGATCCAGGTTACTATGTGTGACAACTGTACTAATGTGTGACTGTGAACCG 4134  
 QY 1837 CysGlnHisGlnSerValCyThrArgLYsProSerAlaProHisGlyLYsThrCyGln 1856  
 Db 4135 TGTGAGCACCAAGTGTGTGTATCCCGCAAGCCAGTGTCCCCCAATGTGCTAATACCTGCAG 4194  
 QY 1857 CysProProAsnTrpLeuGlyProLYsCyAspGlyIleThrArgIleAspGlnProCyAspProArg 1876  
 Db 4195 TGTCCCCCAATTACTTGGGCACTTGTGACATCTGTAGAACAGAGATTGACAGACCTTGTCCCGT 4254  
 QY 1877 GlyTrpTrpGlyHisProThrCyAspGlyProCyAsnCyAspValSerLYsGlyPheAsp 1896  
 Db 4255 GGCTGTGTGGGACATCCCACTATGTGGCCATGTGCAACTGTGATGTGACAAAGGCTTTGAC 4314  
 QY 1897 ProAspCyAsnLYsThrSerGlyGlyLYsCyHisGlyLYsGlyValAsnHisLYsArgProPro 1916  
 Db 4315 CCAGACTGTCAACAAAGACAAAGCGGAGTGTCCACTGTGCAAGAGAAACCACTACCGGCCCCA 4374  
 QY 1917 GlySerProThrCyAspLeuCyAspCyAspTrpProThrGlySerLeuSerArgValCyAs 1936  
 Db 4375 GGCAAGCCCACTGTCTTGTGTGACTGTGACTTACCCCAAGGCTCTTGTCCAGAGTCTGT 4434  
 QY 1937 AspProGluAspGlyGlnCyAspProCyAspAspProGlyValIleGlyArgGlnCyAspArg 1956  
 Db 4435 GACCTTAGAGATGGCCAGTGTCCATGTGCAAGCCAGGTGTCAATCGGCGGTCACTGTGACCGC 4494  
 QY 1957 CyAspAsnProPheAlaGlyValIleThrArgGlyCyGlnValAsnTrpAspSerCyAs 1976  
 Db 4495 TGTACAAACCTTGTGTGTGAGTGCACCAACATGGCTGTGAAAGTAAATTAAAGACGCTGC 4554  
 QY 1977 ProArgAlaIleGlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAlaIle 1996  
 Db 4555 CCAAGACGATTAAGCTGGATCTGTGTGGCCCTGTAACCGCTTGGGCTGTCTGTCT 4614  
 QY 1997 AlaProCyAspProLYsGlySerPheGlyLYsThrAlaValArgHisCyAspGlyHisArgGly 2016  
 Db 4615 GCTCTCTGTCCCAAAAGGCTCTTGTGGACTGTGTGCGCACTGTGATGACACAGGGGG 4674  
 QY 2017 TrpLeuProProAsnLeuPheAsnCyThrSerIleThrPheSerGlyIleuLYsGlyPhe 2036  
 Db 4675 TGGCTCCCCCAACCTCTTCACTGACGTCACATCACTTCTCAAGACTGAAGAGGGCTTC 4734  
 QY 2037 AlaGluArgLeuGlnAlaGlnAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAla 2056  
 Db 4735 GCTAAGGCGCTACAGCGGAATGACTCAGGCTTACACTCAGGGCCCTCCAGCACTAGCC 4794  
 QY 2057 LeuLeuLeuArgAsnAlaThrGlnHisIleThrAlaGlyLYsTrpPheGlySerAspValLYsVal 2076  
 Db 4795 CTGCTCTGTGGCAACGCGCACGAGCACACAGCTGTGCTAATTCCGACGACGTCAAGGTG 4854  
 QY 2077 AlaTrpGlnLeuAlaThrArgLeuLeuAlaHisGlyLeuSerThrGlnArgGlyPheGlyLeu 2096  
 Db 4855 GCTTACCAAGCTGGCGACGGCGGTGTGGGCCACAGAGGACCCAGCGGGGCTTGTGGCTG 4914  
 QY 2097 SerAlaThrGlnAspValHisPheThrGlyAsnLeuLeuArgValGlySerAlaLeuLeu 2116  
 Db 4915 TCTGCCACACAGACGATGACTTCACTGAGAAATGTGTGGGGGTGGGCGCCCTCTG 4974  
 QY 2117 AspThrAlaAsnLYsArgHisLYsTrpGlyLeuIleGlnGlnThrGlnGlyGlyLYsThrAlaTrp 2136  
 Db 4975 GACACAGCCAAACAGCGGCACTGGAGCTGATCCAGACAGACAGAGGGTGTGACCGGCTGG 5034  
 QY 2137 LeuLeuGlnHisLYsTrpGlyValIleTrpAlaSerAlaLeuAlaGlnAsnMetArgHisIleThr 2156  
 Db 5035 CTGCTTCAAGCATTAAGGCTTACGCGATGCGCTGTGGCCCAAGACATGGGCACTTAC 5094  
 QY 2157 LeuSerProPheThrIleValIleThrProAsnIleValIleSerValIleArgLeuAspLYs 2176  
 Db 5095 CTAAAGCCCTTTCACACATGTGTGACGCGCCCAATTTGTATCTCCGTAAGTGGCGTTGTGACAA 5154  
 QY 2177 GlnAsnPheAlaGlyAlaLYsLeuProArgTrpGlyIleAlaLeuArgGlyGlnGlnProPro 2196

Db 5155 GGGAACTTGTGGGGGCGCAAGCTGCCCCGCTACGAGGCGCTGGGGAGAGCCCCCG 5214  
Qy Aspleu1unThrThrVal11leuProgluSerVal1PheArg1unThrProProVal1 2216  
Db 5215 GACCTTAGACAAAGCAATTCCTGCTGAGTCTGTCTTCAGAGAGAGCCCCCGGTGTC 5274  
Qy ArgPro1aGlyProG1yGlu1aGln1upProglu1uLeu1aAArg1aGlnArg 2236  
Db 5275 AGGCCCCAGGCCCCGAGAGAGGCCAGAGAGAGCTGGCAAGCCGCAAGCAGCG 5334  
Qy HisProgluLeuSerGln1yGlu1aVal1aSerVal11le1yTrArgThrLeu1a 2256  
Db 5335 CACCCGAGGCTGAGCCAGGGGTGAGGCTGTGGCCAGCTCATCTACCGCACTCGGCC 5394  
Qy G1yLeuLeuProH1saen1yTrAspProAsp1yAspSer1euaVal1Pro1yAspPro 2276  
Db 5395 GGGGTACTGCTCTCAATCAATAGACCTGACAAAGCGAGCTTGAGAGTCCCAACGGCCG 5454  
Qy 11e11eAsnThrProVal1aSer11eSerVal1H1saAspArg1uGluLeuLeuProArg 2296  
Db 5455 ATCATCAACACACCCGTGTGTAGCATCAGCTCCATGATATGAGAGCTTCTGCCCGG 5514  
Qy 2297 A1aLeuAsp1yBProVal1ThrVal1GlnPheArg1eueuG1unThrg1uaArgThrLys 2216  
Db 5515 GCGCTGGAACAACCGCTCAAGGTGCAAGTCCGCTGCGAGACAGAGAGCGGACCAAG 5574  
Qy Pro11eAsnThrPhe1TrAsnH1Ser11eLeuVal1Ser1yThrG1yG1yTrpSer1a 2336  
Db 5575 CCAATCTGTGTCTTGAGACCAATTCATCTGTGTGAGTGCACAGGTGTGTGTGCGCC 5634  
Qy ArgG1yCyBgluVal1aPheArgAsnG1uSerH1aVal1SerCyBgluCyAsnH1aMet 2356  
Db 5635 AGAGGCTGTGAAGTGTCTTCGCAATGAGAGCCACGTGACCTCCAGTGCACACAG 5694  
Qy 2357 ThSerPheal1aVal1leuMetAspVal1SerArg1uGluAsnG1yG1u11eLeuProLeu 2376  
Db 5695 ACGAGCTTGTGCTGCTCATGAGAGCTTCTGCGGAGATGAGAGATCTTCCCACTG 5754  
Qy 1yAsnThrThrTrVal1a1aLeuG1yVal1ThrLeu1a1a1eLeuLeuThrPhePhe 2396  
Db 5755 AAGACATGACAAATGAGTGTCTAGGTCTACCTTGCTGCTGCTTCTGCTCACTTCTTC 5814  
Qy 2397 PheLeuThrLeuLeuArg11eLeuArgSerAsnGlnH1aG1y11eArgArgAsnLeuThr 2416  
Db 5815 TTCTCACTCTTGTGCGATCTCTGCTCAACCAACAGCGCATCCGATCACTGCA 5874  
Qy 2417 A1a1a1aLeuG1yLeu1aGlnLeuVal1PheLeuLeuG1y11eAsnGln1aAspLeuPro 2436  
Db 5875 GCTGCTGCGGCTGAGCTGAGCTGCTGTCTTCTCTGCGAATCAACAGGCTGACCTCCCT 5934  
Qy 2437 Phe1a1aCyBThrVal11e1a11eLeuLeuH1aPheLeu1yTrLeuCyBThrPheSerTrp 2456  
Db 5935 TTTCCTGCAAGATCATGCTGCTGCACTTCTGTAACCTCTGCACTTCTTCCTGG 5994  
Qy 2457 A1aLeuLeuGln1a1eLeuH1aLeu1yTrArg1a1eLeuThrg1uVal1ArgAspVal1AsnThr 2476  
Db 5995 GCTCTGCTGAGAGGCTTGCACCTGTACCGGGCACTCACTGAGGTGGCGGATGTCAACCC 6054  
Qy 2477 GlyProMetArgPhe1yTr1yTrMetLeuG1yTrpG1yVal1Pro1a1aPhe11eThG1yLeu 2496  
Db 6055 GGGCCCAAGCGCTTCTACTACATGCTGTGGCTGGGGCGGTGCTGCTCATCAACGGGCTA 6114  
Qy 2497 A1aVal1G1yLeuAspProG1uG1yTrG1yAsnProAspPheCyBTrpLeuSer11eTr 2516  
Db 6115 GCGGTGGGCTGTGAGACCCGAGAGGCTACCGGAACCTGTACTTCTGCTGCTCTCATCAT 6174  
Qy 2517 AsnThrLeu11eTrpSerPheal1aG1yProVal1aPheal1aVal1SerMetSerVal1Phe 2536  
Db 6175 GACAGGCTCATCTGAGTGTGTGCTGGCCGGGTGGCTTGTCCGCTCATAGTGTCTTC 6234  
Qy 2537 LeuTr11eLeu1a1a1aArg1aSerCyB1a1a1aArgG1nG1yPheG1u1yB1yB 2556

Db 6235 CTGTACATCTGTGGCGGCGCGGCTCTGTGCTGCCAAGCGGAGGCTTTGAGAGAA 6294  
Qy 2557 GlyProVal1SerG1yLeuGlnProSerPheal1aVal1leuLeuLeuSer1aThrTrp 2576  
Db 6295 GGTCTGTCTGGGCTGAGGCTGCTCTCTGTGCGCTCTCTGTGTGTGAGCGCAAGTGG 6354  
Qy 2577 LeuLeu1a1eLeuLeuSerVal1aSerAspThrLeuLeuPheH1aTr1yLeuPheal1aThr 2596  
Db 6355 CTGTGCGCATGTGCTCTCTCAACAGCAACCTCTCTTCCATCACTCTTGTCTTAC 6414  
Qy 2597 CyAsnCyB11eGlnG1yProPhe11ePheLeuSerTrVal1a1eLeuSer1yG1uVal1 2616  
Db 6415 TGCATATGATCCAGGGCCCTTCACTTCTCTCTATGTGTGTGTGTGCAAGAGGTGC 6474  
Qy 2617 Arg1yVal1a1eLeu1yLeu1a1aCyBSerArg1yBProSerProAspPro1a1eLeuThrThr 2636  
Db 6475 CGGAAAGCATCAAGTGTCTGCAAGCCGAGCCAGCCGAGCCCTGACCTCTGACACCC 6534  
Qy 2637 1ySerThrLeuThrSerSerTrAsnCyBProSerProTr1a1aAspG1yArgLeuTr 2656  
Db 6535 AAGTCAACCTGACCTGTCTTCAACTGCCCCAGCCCTTACGAGATGGGGCGGTGAC 6594  
Qy 2657 GlnProTrG1yAspSer1aG1ySer1eLeuH1aSerThrSerArgSerG1y1yBSerGln 2676  
Db 6595 CAGCCCTTACGAGACTCGGCGGCTCTCTGCAAGACACAGTCTCGGGCAAGAGTCAAG 6654  
Qy 2677 ProSerTr11eProPheLeuLeuArgG1uG1uSer1a1eLeuAsnProG1yGlnG1yPro 2696  
Db 6655 CCAAGCTACATCCCTCTTGTGAGAGAGTCCGCACTGAACCTTGGCCAAAGGAGCC 6714  
Qy 2697 ProG1yLeuG1yAspProG1ySerLeuPheLeuG1uG1yGln1aAspGlnH1aAspPro 2716  
Db 6715 CCGGCTGTGGGGATCAAGGAGCGCTGTCCTGAAAGTCAAGACAGCAGCATGATCTCT 6774  
Qy 2717 AsnThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerG1ySerTr1a1aSer 2736  
Db 6775 GACACGAGTCTCCGACAGTACCTGTCTTAGAAGACAGACAGGTGCTCTTAGTCTCT 6834  
Qy 2737 ThH1aSerSerAspSerG1uG1uGlnG1uGlnG1uGlnG1u1a1aPhePro 2756  
Db 6835 ACCCATCTCATCAGCATGTAAGAGAGAGAGAGAGAGAGAGAGAGGCGGCTTCCCT 6894  
Qy 2757 GlyGlnGlnG1yTrAspSerLeuLeuG1yProG1yVal1GlnArgLeuProLeuH1aSer 2776  
Db 6895 GAGAGACAGGCGGTGATAGCTGCTGGGGCTCGAGAGAGAGTGGCCCTGCAAGT 6954  
Qy 2777 ThPro1yAspG1yG1yProG1y1yTrVal1aProTrpProG1yAspPheG1yThr 2796  
Db 6955 ACTCCAGAGATGGGGCCCAAGGCTGTGGCAAGGCCCCCTGGCCAGAGACTTGGGAGCC 7014  
Qy 2797 Thr1a1yG1ySerSerG1yAsnG1yVal1aProGlnG1uArgLeuArgG1uAsnG1yAsp 2816  
Db 7015 AAGAGAAAGAGATGTGTGCAACCGGGCTCTGAGAGAGCGGTGGGAGATGAGAT 7074  
Qy 2817 A1aLeuSerArgG1ySerLeuG1yProLeuProG1ySerSer1aGlnProH1aB1yB 2836  
Db 7075 GCGCTGTCTGAGAGGGGTCCCTAGGCGCCCTTCCAGGCTCTTCTCCAGCTCAAA 7134  
Qy 2837 Gly11eLeu1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB1yB 2856  
Db 7135 GGCATCTTAGAAGAGATGTGTGCAACCATAGAGAGAGAGAGAGCTCTGCGGCTC 7194  
Qy 2857 ProLeuGlnG1yCyBThrg1ySerSerArg1ySerSer1aSerG1uG1ySerArg1y 2876  
Db 7195 CCGCTGAGAGAA1yGACAGAGGTCTTCGCGGGCTCTCGCTAGTGAAGGCAAGCGGGGC 7254  
Qy 2877 GlyProProProArgProProProProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGln 2896  
Db 7255 GGGCCCCCTCCCGCCACCGCCCGAGAGCTCTCAAGAGAGAGTCAAGAGCGGGCTCAG 7314  
Qy 2897 Pro11e1a1eSer11e1yB1a1aG1yThrVal1aAspG1yAspSerSerG1ySerG1uPhe 2916  
Db 7315 CCAATGCCATGAGCATCAAGAGGACCGGTGATGAGGACTGTCAAGCTCCGAATTT 7374

Qy	2917	LeupPhepaSnpheLeuHts 2923
Db	7375	CTCTTCTTAATTCTGCAT 7395
RESULT 8		
AAID	AA06332	
AC	AA06332	standard; cDNA; 9045 BP.
XX	AA06332:	
XX	26-SEP-2001	(first entry)
XX	DNA encoding seven-pass transmembrane receptor-like protein, MEM1.	
DE		
KW	Seven-pass transmembrane receptor; MEM1; therapeutic; diagnostic; MEM2; human; Alzheimer's disease; Parkinson's disease; cancer; nephrology; female reproductive health; lung disorder; brain disorder; schizophrenia; heart disorder; arhythmia; muscular disorder; clotting deficiency; MEM3; cocalamine deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6; vision-related disorder; neoplastic pathology; MEM7; MEM8; ss.	
XX		
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
XX	CDS	1..9045
XX		/*tag= a
XX		/product= "Seven-pass transmembrane receptor-like protein, MEM1"
XX	WO200144473-A2.	
XX	21-JUN-2001.	
XX	14-DEC-2000; 2000MC-US33909.	
XX	14-DEC-1999; 99US-0170564.	
XX	27-DEC-1999; 99US-0173165.	
XX	27-DEC-1999; 99US-0173362.	
XX	29-DEC-1999; 99US-0173544.	
XX	04-JAN-2000; 2000US-9966564.	
XX	09-AUG-2000; 2000US-0223529.	
XX	13-DEC-2000; 2000US-9966565.	
XX	(CURA-) CURAGEN CORP.	
PA	Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;	
PI	WPI; 2001-398154/42.	
DR	P-PSDB; AAU02196.	
XX		
XX	Novel polypeptide comprising members of protein families (e.g., seven-pass transmembrane receptor proteins) according to presence of domains and sequence relatedness are useful for treating or preventing, e.g., Alzheimer's and Parkinson's -	
XX	Claim 9; Fig 1; 16zpp; English.	
XX		
XX	The sequence represents the coding sequence of seven-pass transmembrane receptor-like protein, MEM1, selected from a group (MEM1-MEM8) comprising members of protein families according to the presence of domains and sequence relatedness, e.g., seven-pass transmembrane receptor protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).	
XX	The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are all useful for treating or preventing a pathology associated with (I) comprising administering (I), (II), or (III) to a subject (preferably a human). In addition, (I), (II), and (III) may be used to manufacture a medicament for treating a syndrome associated with a human disease that is associated with (I). Furthermore, (I) may be used to identify agents that bind to it, screen modulators of its activity and determine the presence or predisposition to a disease associated with altered levels of (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease.	

[illegible]



```

Db      724 AGCTGAAAGTTCCAGATCCAGACTACAGAGTGGCTGTGGAGAACAGACCGGCGGC
Qy      198 ThrProValAlaSerLeuArgAlaIleAspProAspGluGlyValAlaGlyArgLeuGlu
Db      784 ACCCTCATCTCCAGCTGACCGCGCACTACCATGACGGGCGAGGAGGAGCGCGGAGC
Qy      218 TyrThrMetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProVal
Db      844 TATTATCATGAGAGGGCTGTTCGACGAGCGCTCCCGGGGCACTTCGGAATTCAGCTGTGC
Qy      238 ThrGlyAlaValThrThrAlaGluGluLeuAspArgGluThrLysSerThrIleValPhe
Db      904 ACCGGGCGCGTGGACGACGACGCTGACCGCGAGACCAAGAGACGACGCTCTC
Qy      258 ArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle
Db      964 AGGGTAAAGCTCGTGACTACGATACCGCGCGCTCGGACCACTACCTACCTGTC
Qy      278 LeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGluLysArgLeuSer
Db      1024 TTGCTCAAAAGACCAACAGACCAAGCCCGGTCTTGAGAGATCGGAGTACCGGAGCGC
Qy      298 LeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAsp
Db      1084 GTGCGGAGAACTGAGAGGTGGCTACGAGGTGCTGACCATCCGCGCAGACCGCGAC
Qy      318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysProSer
Db      1144 TCGCCATCAACCGCACTTCGCTTACCGGTGTGGGGCGCGCG-----
Qy      338 GluValPheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArg
Db      1192 GAGCTTTCAGCTCAACGAGAGCTGTGGGTGTGAGACACAGGGCGGTGTGAGACCGG
Qy      358 GluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGly
Db      1252 GAGGAGCGCGCCAGATCAAGCTCTGTGTGAGGCAACAGACGAGGGGCGCAATCCGGGC
Qy      378 ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsnAspAsnAlaPro
Db      1312 CCGCTCAAGTCCAGCGCACCGCTGTATCATCGAGTGAAGAGACAGAAACACATACCCC
Qy      398 GlnPheSerGlyIleArgTyrValValGlnValArgGluAspValThrProGlyAlaPro
Db      1372 CAGTTCAAGCAGAGACATACTGCTGTCCAGGTGCCGAGGACGTGGGGCTCAACACGGCT
Qy      418 ValLeuArgValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSer
Db      1432 GTGCTCGAGTGCAGGCGCACGAGCCGAGGACCAAGGCGCAAGACCGGCACTTCATCAGC
Qy      438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp
Db      1492 ATCTCAAGCGGAACTGGCGCGGCGCACTTACTTCATCGCTGAGCGGATCTCGGAT
Qy      458 ValValSerProLeuAspTyrGluThrThrLysGluTyrThrLeuArgValArgAlaGln
Db      1552 GTGATCAACCCCTTGATTTGAGGATGTCCAGAAATACCTCGCTGAGCATTAAGGCCAG
Qy      478 AspGlyValArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp
Db      1612 GAGTGGGGCGCGCGCGCTCATCTTCAGGGGTGTGTCTGTGACGAGTGTGAGT
Qy      498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer
Db      1672 GTCAAGACACACAGCTATCTTTGAGACAGCCCTTCAGAGCCACGAGTGTGAGAAAT
Qy      518 ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn
Db      1732 GTGCGCTGTGGCTTACCCGTGTGCAATTCAGCGGTGACCGGAGCTCTGAGAGAAAC
Qy      538 AlaArgLeuGluIleValArgLeuAla-----GlyValGlyHis-----

```

```

Db      1792 GCCCGGCTGCATATCGCTGTGTGACAGCGGCTTCACCTTTCTGGGGGGCGGACGGCT
Qy      550 -----AspPheProPheThrIleAsnAsnGlyTyrGly
Db      1852 GGGCTTAAGAAATCCCGCCCAACCCCTGACTTCCCTTCAGATTCACAAACGCTCCGCT
Qy      561 TrpIleSerValAlaAlaGluLeuAspArgGluValAlaPheTyrSerPheGlyVal
Db      1912 TGATATCAAGTGTGTGCGGAGCTGAGACCGGAGGAGGTGAGACATCAGCTTCGGGGTG
Qy      581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal
Db      1972 GAGGGGTGAGACCAAGGCTCCCGCCCAAGCTCTTCACACAGGTGTTCATCAGGTG
Qy      601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn
Db      2032 CTGAGCGTAATGACACAGACCCGCTGTTCACGAGCCCACTACAGACTTCGCTGAAAT
Qy      621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis
Db      2092 GAGGATCGCGCGGTGGGAGAGCGGTGTGACCTTGACGAGCCGCGACCGTGAAGCCAAAC
Qy      641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer
Db      2152 AGTGTATTACTTACCACTCAAGCGCGCAACCCGGAACCGCTTGTGACCTCAGAGC
Qy      661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluValGln
Db      2212 CAGAGAGGGGGCGGCTCATACCTTGGCGCTTACTCTGAGCTACAGACAGAGACGAG
Qy      681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal
Db      2272 TACGTGTGGCGGTGACAGATCCGACGAGCAGCGGTGCACATCGCGCATGCTTAATC
Qy      701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn
Db      2332 AACGTCACTGAATGCCAACACCAAGCGCTGTCTTTCAGACTCCCATTAACAGGTGAT
Qy      721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp
Db      2392 GTCACTGAGACAGCGCTGTGGGCACTTCATGCTACCTCAGTGCACAGATGAGGAC
Qy      741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle
Db      2452 ACAGGAGGAATGCCCGCATCACTACGTGATTCAGAGACCCCGTCCGCAAGTCCGCAAT
Qy      761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer
Db      2512 GACCCCGACAGTGGACCATATACACATGATGAGCTGACATATGAGAACCAAGTCCGC
Qy      781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnLysSerAspThrTyr
Db      2572 TACAGCTCAACATCATATGCGCAGAGACAGCGGATCCCGGAAATCAGACACCAACC
Qy      801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSerTyr
Db      2632 CTAGAGATCTCATCTCTCGAGTGCACATGACATGACACCCAGTTCCTGTGGAAATTCAC
Qy      821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr
Db      2692 CAGGGTTCATCTTGTAGAGAGCTTCACCTTCAGACACACATCTTCAGAGTCTTCGACG
Qy      841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGly
Db      2752 GACCGGAGCTCAGAGTCCCAATGGGGGTGTGTGTATCACTTCAGAGGTGGGAGACAGCGG
Qy      861 AspGlyAspPheIleValGluSerThrSerGlyIleValAlaGlnThrLeuArgArgLeuAsp
Db      2812 GATGGGAGCTTCAATCATGAGACCAAGTGTGTGATTCACCAAGCGCGGTGTGAGC
Qy      881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetPro---
Db      2872 CCGAGAAATGTGGCGGTATCAACCTTGGGCTGTGCTGTGAGTCCGGGCGAGTCCCACT

```



QY	900	ProAlaArgThrPrometGluValThrValThrValLeuAspValAsnAspAsnProPro	919
Db	2932	CCCCCTTACCGCCTCCGTTAGAAATCCAGAGTACCATCTTGGAATATGACAAATGCCCC	2991
QY	920	ValPheGluGlnAspGluPheAspValPheValGluGlnAsnSerProIleGlyLeuAla	939
Db	2992	ATGTTTGAAGAAGAACTGGAAGCTGTTTGTGTGAGGAACAAACCCAGTGGGTCGGTG	3051
QY	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetCysGln	959
Db	3052	GTCGCAAAAGATTCCGTCCTAAGACCCCTGATGAAAGCCCTAAGGCCAATCATGATACG	3111
QY	960	IleValGluGlyAsnIlePheProGluValPheGluLeuAspIlePheSerGlyLeuLeuThr	979
Db	3112	ATTGTGGAAAGGGGACATCCGCGCATTTCTTCCAGCTGGACCTGCTCAAGGGGACCTGGCT	3172
QY	980	AlaLeuValAlaPheLeuAspTryGluAspArgProGluValThrValIleGlnAlaThr	999
Db	3172	GCCATGCTGGAGACTGCACTTGAGGTCGGCGGGAGTATGTCGTGTGTGTGACGCCACG	3231
QY	1000	SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn	1019
Db	3232	TCGGCTCCGCTGCTGATGCGCGACCCAGGTGCAACATCTTCTGTGACACAGATATCAAC	3291
QY	1020	ProProValIleGluGlyAsnPheGluIleLeuPheAsnAntyrValThrAsnArgSerSer	1039
Db	3292	CCGCTCTGCTGCGCCGACCTTCAGATCTCTTCCAAACAATGTGCACAAACAGTCCAC	3351
QY	1040	SerPheProGluGlyAlaIleGlyValArgValProAlaHisAspProAspIleSerAspSer	1055
Db	3352	AGTTTCCCCACCGCGGTATGCGCTGCATCCCGGCCATGACCCCGAGTGTGCACACGC	3411
QY	1060	LeuThrTyrSerPheGluArgGlyValAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr	1079
Db	3412	CTCAACTACACCTTGCTGTGACAGGGCAACAGAGCTGGCTGTGCTGTGACCCCGCACG	3471
QY	1080	GlyGluLeuLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaIleMetSer	1099
Db	3472	GCGCAACTGCAGCTCAGCCGCGACCTGTGACAAACACCGCGCTGGAGCGCTCAATGAG	3531
QY	1100	ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle	1119
Db	3532	GTCGTCGTGTGTGATGAGCATTCACAGCTTCACGCGCTTCTGCACTTCGCTGTACCATC	3591
QY	1120	IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu	1139
Db	3592	ATCACGCGACGACATGCTGACCAACAGCATCACTGTCGCCCTGGAGAAATGTCACAGAG	3651
QY	1140	ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr	1155
Db	3652	AAAGTTCCTGCTCCCGCTGCTGCGCTCTTTCGTGAGGGGGTGGCGCGCTGCTGCCAC	3711
QY	1160	ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis	1179
Db	3712	ACCAAGACGACGCTTCTGCTTCACCTTCAGAACGACACCGACGTC---AGCTCCAAC	3766
QY	1180	IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPhe	1199
Db	3769	ATCTCGAAGCTGACCTTCTCGCGCGCTGCTGCGCTGCGCGCGCTCCCGGC-----CAGTTC	3822
QY	1200	LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle	1219
Db	3823	TTCCTGCTCGAGACCTCAGAGACGAGATCTACCTGAATCGGACGCTGTACCAACCATC	3882
QY	1220	SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn	1239
Db	3883	TCCACGCGACGCGTCTGCTCCCTTCGACGACAAACATCTGCTGCGGAGACCTCGGAGAAC	3942
QY	1240	TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer	1259
Db	3943	TACATGAAGTGCAGTCTCGCTTCGGAATTGACACAGCTCCGCGCCCTTCTCAGCTCCAC	4002

QY	1260	SerValLeuPheArgProIleHisProValGlyIleuAspCyValGyrProProGly	1279
Db	4003	AccGGGCTCTCCGGCCCATCAACCCCAACAAGCCCTGCCTGCCGCCGCCGCC	4062
QY	1280	PheThrGlyAspArgYrCySGIuThGlyValAlaPheLeuCyYrSerArgProCyGlyPro	1299
Db	4063	TTCAACCGGCGACTCTGGAGAACCGAGATCAACTCTCTCTACTCCGACCCCTGGCGGCC	4122
QY	1300	HisGlyArgCyValGSerArgGlyGlyIYrThrCyLeuCyValGAspGlyYrThr	1315
Db	4123	AACGCCCTCTCCGACGCCGAGAGGCGGCTCAACCTGCAGTGCCTTGAGAACTTCACT	4182
QY	1320	GlyGlyHisGyrGlyValSerAlaArgSerGlyArgGyrSerProGlyValCyValSer	1333
Db	4183	GGAAGACCTGTAGAGTGGATGCCCGCTCAAGCGCTGTGCCAAGCGGGGTGTGAAAGAC	4242
QY	1340	GlyGlyYrThrCyValAlaMetLeuLeuValGlyIYrPheLysCyAspCyProSerGlyAsp	1355
Db	4243	GGGGGCACTCTCGTGAACCTGCTCATCGCGGCTTCCACTGCGTGTCTCTGGCGAG	4302
QY	1360	PheGlyLysProYrCyGlnValThrThrArgSerPheProAlaHisSerPheIleThr	1379
Db	4303	TATAGAGGCTCTCTGTGAGGTACCAACGAGAGCTTCCGCCCACTCTTGCTGATCC	4362
QY	1380	PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrGlyGly	1399
Db	4363	TTCCGGGGCTCGAAGACAGCGCTTCACTTCAACATCTCCCTCAAGTTTGCACTCAGAA	4422
QY	1400	ArgAspGlyLeuLeuLeuYrYrAsnGlyValArgPheAsnGlyLysHisAspPheValAlaLeu	1419
Db	4423	AGGAACGGCTTGTCTCTCAACAGCGCGCTTCATGAGAAAGACAGCACTTCAAGCCCTG	4482
QY	1420	GlyValIleGlnGlnValGlnLeuThrPheSerAlaGlyGyrSerThrThrThrVal	1439
Db	4483	GAGATCGTGAACGACAGAGTCACTACTCTCTGAGCGAGACAAACAACAGCGTG	4542
QY	1440	SerProPheValProGlyGlyValSerAspGlyGlnTrpHisThrValGlnLeuLysYr	1455
Db	4543	GCACCGAAGGTTCACAGTGGTGAAGTACGGGCGGTGGCACTCTGTGCAGGTGACATAC	4602
QY	1460	YrThrLeuLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnLysVal	1479
Db	4603	TACAACAAGCCCAATATTGTGCACCTCGGCTGTGCCCAATGGCGGTCCGGGAAAAAGTGG	4662
QY	1480	AlaValValThrValAlaAspGlyCyAspThrGlyValAlaLeuArgPheGlySerValLeu	1499
Db	4663	GCGGTGGTGAAGTGGATGTGTGTGACAAACATGCTGTGGCCCTTTGAAAAGACATC	4722
QY	1500	GlyValenYrSerCyValAlaGlnGlyThrGlnGlyGlySerLysLysSerLeuAspLeu	1519
Db	4723	GGGAACTAAGCTCGCTGCCGACGAGGCACTCAAGCCGGCTTCAAGAACTCCCTGATCTG	4782
QY	1520	ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMet	1539
Db	4783	ACCGGCCCTACTCTCTGGGGGGGTGTCCCAACTGCGCAAGAACTTCCAGTGCACAAAC	4842
QY	1540	ArgGlnPheValGlyCyMetArgAlaLeuGlnValAlaPheSerArgHisIleAspMetAla	1555
Db	4843	CGGCAAGTTCGGGCTGCATGCGGAACCTGTCACTGACACGGCAAAAATGTGACATGCC	4902
QY	1560	AspPheIleAlaAsnAsnGlyThrValProGlyCyProAlaLysValValAspValCyAsp	1579
Db	4903	GGATTCATCGCAACATGGCAACCGGAAAGCTGCCTGTGGAAGAACTTCTGCAT	4962
QY	1580	SerAsnThrCyHisIleAsnGlyGlyThrCyValAlaAsnGlnTrpAspAlaHisSerCyGlyL	1599
Db	4963	GGAGCGCGGTGTCAAGATGAGGAGCACTGTGTCAAGAGTGAATATGTATCTGTGAG	5022
QY	1600	CysProLeuGlyPheGlyGlyLysSerCyValAlaGlnLeuMetAlaAsnProGlnHisPhe	1619
Db	5023	TGTCACTCTCGAATTCGGCGGAAAGAACTGTGAGAACCAATGCTCAACCCCACTCTTC	5082
QY	1620	LeuLysSerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpYr	1639



QY	2343	PheA	AGANGLSuSEHivAlaSerCySGInCySaenH1smETThSerPheAlaValLeu	2365
Db	7303	TCAGGAACCGGACACATGTCGCCCTGCAGTGCACACAGCCAGCTTTCGGGTGCTC		7367
QY	2363	MetA	APValSerAArgGluAnGlyGluLeuProLeuYsThSerLeuThyVal	2387
Db	7363	ATGATATCTCCAGGCGGTGAGAACGGGGAGGCTCGCTCTTGAAATATTGTACCTAATGCC		7422
QY	2393	AlaLeu	GlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeuLeuArg	2402
Db	7423	GCTGTGCTCTGTGTACCTGCAAGCCCTGCTGTGGCTTCGTCTCTTAAGCCTGTGTCCGC		7482
QY	2403	IleLeuA	ArgSerAAsGlnhiAgLyLLeaArgAseuLeuThrAlaAlaLeuGlyLeuAla	2422
Db	7483	ATGCTGGCCTCCAACTGCAACCTGCAAGCATTCACAAAGCACCTCGCCGTGGCCTCTTCTCTCT		7542
QY	2423	GlnLeuVal	PheLeuLeuGlyYTLaaGlnAlaAseuProPheAlaCySerThValIle	2442
Db	7543	CAGCTGTGTTTCGTGATTTGGATTCACACAGACGGAAAAACCGTTTCTGTGCACAGTGGTT		7602
QY	2443	AlaIleLeu	LeuHisPheLeuThyLeuCySerThPheSerTPAlaLeuLeuGlnAlaLeu	2466
Db	7603	GCATCTCTCTCCACTAATCACTCAATGACACCTTGTGCTGACCTCTGTGAGAGCTGTG		7662
QY	2463	HisLeuTy	ArgAlaLeuLeuThrGluValAlaArgAseValaenThGlyProMetAArgPheTy	2482
Db	7663	CATGCTACCCGACATGCTACCGAGGTGGCAACATCGAACGGGGCCCATGGGTTCATAC		7722
QY	2483	TyrMetLeu	GlyYTPGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAsePro	2502
Db	7723	TACGTCGTGGGCTGGGGATCCCCGGCCATGTGCACAGACATGGCGGTGGCTGGACCC		7782
QY	2503	GluGlyTy	ArgYAsnProAsePheCyStrPheuSerIleTyYAspThrLeuIleTrpSer	2522
Db	7783	CAGGGCTACGGGAACCCGACCTTCGTGCTGTGCTTCACAGACACCTTAATTGGAGC		7842
QY	2523	PheAla	GlyProValAlaPheAlaValaSerMetSerValPheLeuTyIleLeuAlaAla	2542
Db	7843	TTTCCGGGGCCCATCGAGCTGTATATATCATCAACACAGTCACCTTGTCTTATCTGCA		7902
QY	2543	ArgAla	SeCyAlaIlaGlnArgGlnGlyPheGlySuLySGlyProValSerGlyLeu	2562
Db	7903	AAGCTTCTCTCCAAAGAAAGACCACTATTATGAGAAAAAGGATGTCTCCCTGCTG		7962
QY	2563	GlnPro	SerPheAlaValLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSer	2582
Db	7963	AGGACGCAATTCCTCTGCTGCTGTCTGCTACAGCGCACCTGGCTGTGGGCTGTGGCT		8022
QY	2583	Vala	MetSerAAPTnLeuLeuPheHisTyIleLeuPheAlaThrCyAseuCyIleGlnGly	2602
Db	8023	GTGAACCGGACATGCACTGAGCTTTCATCACTCTTCCGACATTCACGGGCTTTCACAGGC		8082
QY	2603	ProPheIle	PheLeuSerTyValValaLeuSerTyGluValAArgLySaAlaLeuLyS---	2621
Db	8083	CCCTTTCGTCTCTTTTCCACTGCTGTCTCAACAGAGAGTCCGGAAGCACTTAAGGC		8142
QY	2622	LeuAla	CySerAArgLyPProSerProAseProAlaLeuThrThyLeuSerThr--Leu	2640
Db	8143	GTGCTCGCGGAGAGAGACTCACCTGAGAGACTCCGCACACACAGGGCCACCTGCTG		8202
QY	2641	Thi	SeSerTyYAsnCyPProSerProTyZnAlaapGly-----ArgLeuTyGlnPro	2658
Db	8203	ACGGGCTCTCCATCTGACACACACCTTCGGTGAAGGGCCTGACATGCTGCACAGAC		8262
QY	2659	TyrGlyA	SeSerAlaGlySerLeuHisSerThiSerAArgSer-----	2672
Db	8263	TTGGGCGAGTCCACCGCTCGCTGGAAGCATCTGTCAGAGATGAAGGATCCAGAAAGTCC		8322
QY	2673	-----Gly	YsSerGlnProSerTyIlePro	2681
Db	8323	GGCGTATCTCTGGGCTGTGAAGGGGACGCCACGGAGAGCCAGACGGCTCTCTAATGCC		8382

D	b		8383	AGGAGCTGCAGAGAT-----	CCCCCTGGC-----	8406							
Oy			2682	PneumoniaeAgglutinseraIaleAuaanProGIyngInGIyProPGIyleuGIyaSp		2701							
Oy			2702	ProGIyserieuPheleuGIyGLInaSpGIInHIsaSPProaSPThraSPseraSP		2721							
D	b		8407	-----	CAAGATTCCGACTCGAGT	8424							
Oy			2722	SerapSeuSerLeuGIuaAsPaSPlnSerGIySerTYraIasErThrHisSerSerasp		2741							
D	b		8425	AGCAGCTGTCTCTG---GATGACAGAGCACGCTTTACGCTCTCTCACACTTCGTCAAC		8488							
Oy			2742	SerGIyngInGIyngInGIyngInGIyngInGLuaIaaIaapeProGIyGLInGIyTrp		2761							
D	b		8482	AGCAGAGACATGGGCTGGAGCTGAGGAATA-----	TGG	851							
Oy			2762	AspSerLeuLeuGIyProGIyalaGLuaRgLauProLewHisSerThrProLy---Asp		2788							
D	b		8518	GACCCGGCCAGGGGC-----GCCATCCACAGACACCCCAAAGGGGAC		8559							
Oy			2781	GIyGIyProGIyProGIylylaaIaProTPProGIylaSPPhelGIYThrThralaIyGLu		2800							
D	b		8560	GCTGTGGCCAACAACGTTCCGGCCGGCTGGCCCAACAGACCCTGGCTGAGAGTGAAGT		86119							
Oy			2801	SerSerGIyangeIyalaPro-----		2807							
D	b		8620	GAGGACCCGACGGCCAGACCCCGCTGAAGGTGAGACAAGTCACGCTGAGCTGCAC		86797							
Oy			2808	---GIyGIyurGLEuaRgIuaAnGIyASpaIalauSerarRgIyngInGIySerLeuGIyPro		2826							
D	b		8680	CAGAGAGAGACAGGGCAGTCACTCGTAGAGTAACCCCGGACACAGAGAGCGGGGGCGCA		87399							
Oy			2827	LeuProGIySerSerIaagInPro-----HisIySGIyleuIyalyIyCyS		2843							
D	b		8740	GCCAGGCTTGCTAGACAGCCAGCCCCAGACAGAGAAGAAAGCATCTTGAAAAATPAAGTC		87999							
Oy			2844	Leu-----ProThrIleSerGIuIySer-----SerLeuLeuARgLeuPro		2857							
D	b		8800	ACCTACCCGCGCGCTGACGCTGACGAGACAGACAGCTGAAGGGCCGCTCCGGAGAGAG		88595							
Oy			2858	IeuGIuGIyCyEThrGIySerSerIaRgIySerSerIaSerGIuGIySerArGIyGIy		2877							
D	b		8860	CTGGCGGACGTGAGAGAGAGCCCAACATCTCCGCGACAGCTTCCCTCGGGCTCGGGCGG		89119							
Oy			2878	Pro-----ProProaRgProProProProRgInSerLeuGIuGIuLeuAn		28933							
D	b		8920	CCCGACTGCCCATCAACATCAAGAGCCCTGGAGGGAGCGGGGGCCCTGACCACTCAAC		89797							
Oy			2894	GIYValMePrroIleaMetSerIleIyalaGIYThrValaSpGIuaAPSerSerGIy		2913							
D	b		8980	GGGGTC-----GCCATGAATGTCGCACTGGAGCGCCCAAGGCCGATGCTCCGAC		9030							
Oy			2914	SerGIu 2915									
D	b		9031	TCTGAG 9036									
<hr/>													
RESULT 9													
AAS59825	ID	AAS59825	standard; cDNA, 11389 BP.										
XX	AC	AAS59825;											
DT	16-JAN-2002	(first entry)											
DX													
XX	Human novel cytokine encoding cDNA 790CIP2C_4 #1.												
KM	Human; ssp; cytokine; cell proliferation; cell differentiation;												
KM	antiinflammatory; stem cell growth factor; activin; inhiblin; cancer;												
KM	nervous system disease; neuropathy; Alzheimer's disease;												
KM	Parkinson's disease; Huntington's disease; spinal cord disorder;												
KM	head trauma; stroke; myeloid cell disorder; lymphoid cell disorder;												
KM	platelet disorder; thrombocytopenia; stem cell disorder;												
KM	epileptic anaemia; tissue regeneration; wound healing; ulcer;												

KW osteoporosis; osteoarthritis; bone degenerative disorder; SCID;  
 KM periodontal disease; fibrosis; reperfusion; immune disorder;  
 KM severe combined immunodeficiency; infection; autoimmune disorder;  
 KM multiple sclerosis; rheumatoid arthritis; diabetes mellitus; allergy;  
 KM asthma; coagulation disorder; haemophilia; sepsis; nephritis;  
 KM inflammatory bowel disease; food supplement.  
 OS Homo sapiens.  
 PN WO200175093-A1.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US10484.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PR 22-SEP-2000; 2000US-0668680.  
 PR 23-OCT-2000; 2000US-0695618.  
 PR 30-NOV-2000; 2000US-0728711.  
 PR 14-MAR-2001; 2000US-0728711.  
 XX  
 PA (HSE-) HYSEQ INC.  
 XX  
 PT Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J, Xu C;  
 PT Yang Y, Zaho QA, Chen R, Wang D, Goodrich RW, Liu C, Drmanac RT;  
 DR WPI: 2001-626432/72.  
 DR P-PSDB; AAU65533.  
 PS  
 PS Claim 1; Page 164-177; 336pp; English.  
 XX  
 CC The invention relates to isolated human polypeptides (which may be  
 CC cytokines) and the polynucleotides encoding them. The protein is useful  
 CC for identifying a compound which binds to it (e.g. modulators, agonists  
 CC and antagonists). The polynucleotides are useful as an array for mismatch  
 CC detection. The proteins and nucleic acids are useful as nutritional  
 CC sources or supplements. The protein exhibits activity relating  
 CC to cytokine, cell proliferation, cell differentiation, antiinflammatory,  
 CC stem cell growth factor activity, immune stimulating or immune  
 CC suppressing and activin or inhibin related activities. The proteins (and  
 CC antibodies raised against them) and nucleic acids are therefore useful in  
 CC the diagnosis and treatment of diseases and disorders such as cancer,  
 CC central and peripheral nervous system diseases and neuropathies,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, spinal cord disorders, head trauma, cerebrovascular  
 CC diseases, stroke, myeloid or lymphoid cell disorders, platelet disorders,  
 CC thrombocytopaenia, stem cell disorders, aplastic anaemia, for  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, and in tissue repair, healing of burns, incisions, ulcers, for  
 CC treating osteoporosis, osteoarthritis, bone degenerative disorders, or  
 CC periodontal disease, lung or liver fibrosis, reperfusion injury in  
 CC various tissues, various immune deficiencies and disorders including  
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,  
 CC autoimmune disorders (e.g. multiple sclerosis, rheumatoid arthritis,  
 CC diabetes mellitus, myasthenia gravis), allergic reactions and conditions,  
 CC such as asthma or other respiratory problems, coagulation disorders,  
 CC haemophilia, septic shock, sepsis, arthritis, nephritis and inflammatory  
 CC bowel disease, viral infection and are useful in altering bodily  
 CC characteristics. The present sequence encodes a novel protein of the  
 CC invention.  
 CC  
 XX  
 SQ Sequence 11389 BP; 2217 A; 3596 C; 3401 G; 2175 T; 0 other;

[illegible]

258 ArgVal1ThrAlaGlnAspHisGlyMetProArgHisSerAlaLeuAlaThrLeuThrIle 277  
Db AGGGGTAAGCCGGAGCTACAGTACGCGCGCTCGGCCACCACTCACTCACTGCT 1023  
278 LeuVal1ThrAspThrAsnAspHisAspProVal1PheGlnGlnGlnIleTyrTyrGlyGlySer 297  
Db TTGGTTCMAAGACACCAACCAACCAACCGGCTTCGACAGTCCGAGTACCGGACGCGC 1083  
298 LeuArgGlnAsnLeuGlnValGlyTyrGlnVal1LeuThrValArgAlaThrAspGlyAsp 317  
Db GCGCGGAGAACTCGAGGTGGCTACGAGGTGCTGACATCCGCGCCAGCGACCGCGAC 1143  
318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlySerGlyGlySerProSer 337  
Db TCGCCCTTCACGCGCACTTGGGTACCGCGTGGGGGCGCGG----- 1191  
338 GluVal1PheGlnIleAspProArgSerGlyVal1IleArgThrArgGlyProValAspArg 357  
Db GAGCTCTTCAGCTCAACGAGCTCTGGCGTGGAGCAACGCGCGGTGAGACCGG 1251  
358 GlnGlnVal1GlnSerTyrGlnLeuThrVal1GlnAlaSerAspGlnGlyArgAspProGly 377  
Db GAGGAGGCGCGCGAGTACAGCTCCGCGTGGAGGCAACGAGGCGGCGCATCCGCGC 1311  
378 ProArgSerThrThrAlaAlaValPheLeuSerVal1GluAspAsnAspAsnAlaPro 397  
Db CGCTCAAGTCCACCGCCACCGCTGTACCTGAGTGGAGGAGCAACCACTCACTCC 1371  
398 GlnPheSerGlnValArgTyrVal1GlnValArgGlnAspVal1ThrProGlyAlaPro 417  
Db CAGTTCAAGGACAGAACTACGAGTCCAGGTCGCGCGAGAGCGTGGGCTCAACGCGCT 1431  
418 ValLeuArgVal1ThrAlaSerAspArgAspGlySerAsnAlaValAlaHisTyrSer 437  
Db GTGCTGCGAGTCAAGCGACCGAGCCGAGCAAGGCGGCGCATTCACATCAAGC 1491  
438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp 457  
Db ATCCCTACCGGGAGACGTGGCGCGGCTTACCTGACTCGCTGAGGCGGATCTCGAT 1551  
458 ValValSerProLeuAspTyrGlnThrThrTyrGlnGlyTyrThrLeuArgValArgAlaGln 477  
Db GTGATCAACCCCTTGGATTCAGAGATGTCGAAATACCTCGCTAGCATTAAGCCGAG 1611  
478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuVal1ThrVal1GlnValLeuAsp 497  
Db GATGGGCGCGCGCGCGCTCATCAATCTTCAGGGGTGGTGTCTGTGAGGTGTGAT 1671  
498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
Db GTCAACGACMACGAGCTTATCTTGTGACAGCCCTTCGAGCGCACGCTGTGAGAGAT 1731  
518 ValProLeuGlyTyrTyrLeuValLeuHisVal1GlnAlaIleAspAlaAspAlaGlyAspAsn 537  
Db GTGCCCTTGGCTTACCGGTGTGACATTCAGGGGTGGAGCGCGGACTCTGGAGAGAAC 1791  
538 AlaArgLeuGlnTyrArgLeuAla-----GlyVal1GlyHis--- 549  
Db GCGCGGCTGACACTATCGCTGTGAGCAACGCGCTCACTTCTGGGGGCGCGAGCGCT 1851  
550 -----AspPheProPheThrIleAsnAsnGlyTyrGly 560  
Db GGGCTTAAGATCTCGCCCAACCCCTGACTTCCCTTCAGATCAACAACGCTCCGCT 1911  
561 TrpIleSerValAlaAlaGlnLeuAspArgGlnGlnVal1AspPheTyrSerPheGlyVal1 580  
Db TGGATCACAGTGTGCGGAGCTGAGCCGCGAGGAGGTGGAGCATTAAGCTTCGGGGTG 1971  
581 GlnAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValSerVal1ThrVal1 600  
Db GAGGCGGTGAGACCGGCTCGCGCCCATGAGCTCTCCACAGGCTGTCCATCAAGCTG 2031  
601 LeuAspVal1AsnAspAsnAspProThrPheThrGlnProGlnTyrThrValArgLeuAsn 620

2032 CTGACGGAATGACACAGACCCGCTGTTCAAGAGCCCACTCAAGCTTCGCTGAT 2091  
Db GlnAspAlaAlaVal1GlyThrSerVal1ThrVal1SerAlaVal1AspArgAspAlaHis 640  
2092 GAGATGCGCGCTGGGGAGAGCGGTGACCTTCGAGGCGCGCGAGCCGAGCGGAC 2151  
641 SerVal1IleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
Db AGTGATATTCATCAAGCTCAAGCGGCGGACCCCGAACCCTTTGACTCAAGCAC 2211  
661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrGlnGlnArgGln 680  
Db CAGAGGCGCGCGCTCATCAACCTGCTGACTTCACTGACTCAAGCAGAGCAGCAG 2271  
681 TyrValLeuAlaVal1ThrAlaSerAspGlyTyrArgGlnIleAspThrAlaGlnIleVal1Val 700  
Db TACGTGCTGGCGGTGACAGATCCGAGCGCACGGTGGCACACTGGCGCATGTCTATTC 2331  
701 AsnVal1ThrAspAlaAsnThrHisArgProVal1PheGlnSerSerHisTyrThrValAsn 720  
Db AACGTCACTGATGCAACACCAACGAGCTGTCTTTCAGAGCTCCCATTAACAGTAGT 2391  
721 ValAsnGlnAspArgProAlaGlyTyrThrVal1ValLeuIleSerAlaThrAspGlnAsp 740  
Db GTCAAGTGGACAGCGCTGTGGGCACTTCATCTCACTCCAGTGGCCACGATGAGGAC 2451  
741 ThrGlnGlnAsnAlaArgIleThrTyrPheMetGlnAspSerIleProGlnPheArgIle 760  
Db ACAGAGAGATGCCCGCATCCTACGATGATTAAGACCCCGGCGGACATTCGCGAT 2511  
761 AspAlaAspThrGlyAlaVal1ThrThrGlnAlaGlnLeuAspTyrGlnAspGlnValSer 780  
Db GACCCCGACAGTGCACCATGATACCATGATGAGGTGAGCTTATGAGAACCAAGTGGCC 2571  
781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db TACAGCTGACCATCAATGAGCGCCAGAGACAGGCAATCCCGAGAAATCAAGACCAACC 2631  
801 LeuGlnIleLeuVal1AsnAspVal1AsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db CTAGAGATCTCATCTCGATCCGATGCCAATGACATACCCAGTTCCTGTGGGATTTCTAC 2691  
821 GlnGlySerVal1TyrGlnAspVal1ProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db CAGGCTTCACTTGTGAGATGCTCCACCTCGACGACATCTCCAGGTCTCGCACG 2751  
841 AspArgAspSerGlyLeuAsnGlyArgVal1PheTyrThrPheGlnGlyGlyAspArgGly 860  
Db GACCGGACCTCAAGTCCCAATGGCGCTGTCTGTACCTTCCAGGGGTGGAGAGACGCGC 2811  
861 AspGlyAspPheIleVal1GlnSerThrSerGlyIleVal1ArgThrLeuArgArgLeuAsp 880  
Db GATGGGACCTTCAATCAAGACCCACCTCCGCTGTGATTTGGACCCGCGGCTGGAC 2871  
881 ArgGlnAsnVal1AlaGlnTyrVal1LeuArgAlaTyrAlaValAspGlyGlyMetPro--- 899  
Db CGGAGAGATGGCGGTGATCAACCTTGGGCTGTGCTGTGATCGGGGAGTCCACT 2931  
900 ProAlaArgThrProMetGlnVal1ThrVal1ThrValLeuAspVal1AsnAspAsnProPro 919  
Db CCCCTTACCGCTCGTGAAGATCAAGGTACCAATCTTGACATTAATGACAAAGCCGCC 2991  
920 ValPheGlnGlnAspGlnPheAspVal1PheVal1GlnGlnAsnSerProIleGlyLeuAla 939  
Db ATGTTTGAAGAGACGAACCTGAGCTTTGTTGAGAGAGAACCAACAGTGGGTGGGTG 3051  
940 ValAlaArgVal1ThrAlaThrAspProAspGlnGlyTyrAsnAlaGlnIleMetTyrGln 959  
Db GTGGCAAGATGTTGCTTACGACCCCTGATGAGAGGCGCTTATGCGCATGATATCAAG 3111  
960 IleVal1GlnGlyAsnIleProGlnVal1PheGlnLeuAspIlePheSerGlyGlnLeuThr 979

Dh 3112 ATTGTGAGAGGAGCATGCGGCAATTTCTTCAGCTGAGCACTGCTCAACGGGAGCACTGCGT 3171  
Qy 980 AAlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
Dh 3172 GCCATGTGAGCTGGAATTTAGAGTCCGGCGGAGATATGCTGTGTGGTGGCAGGCGAG 3231  
Qy 1000 SerAlaProLeuValSerArgAlaThrValIleValArgLeuAspArgAspAsn 1019  
Dh 3232 TCGGCTCCGTGGTGGAGCGAGCAAGCTGCAATCTCTCGTGGACCAATGACAAAC 3291  
Qy 1020 ProProValLeuIleAspPheGluIleLeuPheAsnAspTyrValThrAsnArgSer 1039  
Dh 3292 CCGCGTGTGCGCCGCACTTCAGATCTCTTCAACAACATATGTCACCAACAGTCCAC 3351  
Qy 1040 SerPheProGluGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Dh 3352 AGTTTCCCAACCGGCGTGAATCGGCTGATCCCGGCCATGACCCCAAGTGTGACAGC 3411  
Qy 1060 LeuThrTyrSerPheGluArgGluAsnGluLeuSerValLeuLeuAsnAlaSerThr 1079  
Dh 3412 CTCACACTACACTTCCTGCGAGGCGCAACGAGCTGCGCTGTGCTGCGTGGACCCCGCACG 3471  
Qy 1080 GAGluLeuIleLeuSerArgAlaAlaAspAsnAspArgProLeuGluAlaIleMetSer 1099  
Dh 3472 GCGGAACTGAGCTCAAGCCGCGACCTGGAACAACAACCGCGCTGAGGCGCTCATAGAG 3531  
Qy 1100 ValLeuValSerAspGluValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Dh 3532 GTGTCTGTCTATATGATCACTCAAGGCTCAAGGCTTCGACCCCTGCTGTCAACATC 3591  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Dh 3592 ATACAGGACGACATGTGACCAACAGCATCATCTGCGCTGGAGAACATGTCCTCCAGAG 3651  
Qy 1140 ArgPheLeuSerProLeuLeuGluIlePheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
Dh 3652 AAGTCTCTGTCCTGCTGCGGCTCTCTGAGAGGAGGAGGCGCGCTGCTGTCCACC 3711  
Qy 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGluGlyHis 1179  
Dh 3712 ACCAAGACGACAGCTTCGTCTTCAACATCCAGAACACACCGACGTC--AGTCCAAC 3768  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGluProGluGlyProProPhe 1199  
Dh 3769 ATCTTGAACTGACCTTCTGCGCGCTGCTGCGCGCTGCGCGCGC-----CAAGTTC 3822  
Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuThrAlaIle 1219  
Dh 3823 TTCCTGTCGAGAGCACTGCGAGAGCAATCTAATCTGAATCGAACGCTGTGACCAACATC 3882  
Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Dh 3883 TCACAGCAGCGCGTGCCTTCGACCAACATCTGCGCGAGCGCTGCGAGAAC 3942  
Qy 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Dh 3943 TACATGAAAGTGGCGTCTGCTGCAATGCAAGCTCCGCGCTTCCTCAAGTCCAC 4002  
Qy 1260 SerValLeuPheArgProIleHisProValGlyIleLeuArgCysArgCysProProGly 1279  
Dh 4003 ACCGTGCTCTTCGCGCCATCCACCCATCAACGCGCTGCGCTGCGCTGCGCGCGC 4062  
Qy 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Dh 4063 TTCACCGGCGACTCTGCGAGACGGAAGATGCACTCTGCTACTCCGACCGGTGGCGGCC 4122  
Qy 1300 HisGlyArgCysArgSerArgGluGlyIleTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Dh 4123 AACGCGCGCTGCGAGCGCGAGCGCGCTACACCTGCGAGTCTTCGAGAACCTTCACT 4182  
Qy 1320 GlyIleHisCysGluValSerAlaArgSerGlyArgCysThrProGluValCysLeuAsn 1339  
Dh 4183 GGAAGACATCTGTGAGGTGATGCCGCTCAAGCGCGCTGTGCCAACGGGGGTGTGCAAGAC 4242

Qy 1340 GlyIleThrCysValAlaAsnLeuValGlyIlePheLysCysAspCysProSerGlyAsp 1359  
Dh 4243 GGGGCACTGTGTGTAACCTGTCTACCGCGGCTTCACCTGCGTGTCTCTCCGCGAG 4302  
Qy 1360 PheGluLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Dh 4303 TATGAGAGGCGCTTACTGTGAGGTGACCAACAGAACCTTCGCCAGTCTCTGTACC 4362  
Qy 1380 PheArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAlaThrArgGlu 1399  
Dh 4363 TTCGAGGCTGAGACAGGCGCTTCACTCACTCACTCACTCACTCACTTGGCACTAGAA 4422  
Qy 1400 ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGluLysHisAspPheValAlaLeu 1419  
Dh 4423 AGGAACGCTTGTCTCTCAACAGCGCGCTTCATATGAGAACAGACATTCATCCCGCTG 4482  
Qy 1420 GluValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
Dh 4483 GAGATCTGGAACGACAGGTGCACTCACTTCTGAGGCGGAGAACCAACACGCTG 4542  
Qy 1440 SerProPheValProGluGlyValSerAspGluThrPheIleThrValGlnLeuLysTyr 1459  
Dh 4543 GCACGAAAGTTCCTCACTGTGTGAGTACGAGCGGTGCACTGTGTGAGGTGCACTAC 4602  
Qy 1460 TyrAsnLysProLeuLeuGluGlyIleThrGlyLeuProGlnGlyProSerGluGlnVal 1479  
Dh 4603 TACAACAAGCCCAATATTTGGCCACCTGGGCGTCCCAATGGGCGCTCCGGGAGAAAGATG 4662  
Qy 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaAlaLeuArgPheGlySerValLeu 1499  
Dh 4663 GCGGTGTGACAGTGAATGATGTGACCAACCAATGCGTGTGCGTTGGAAAGACATC 4722  
Qy 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLeuSerLeuAspLeu 1519  
Dh 4723 GGAATCAAGCTGTGCGCTGCCAGGCACTCAAGCCGCTCCAAAGATCCTGTGATCG 4782  
Qy 1520 ThrGlyProLeuLeuLeuGlyIleValProAspLeuProGluSerPheProValArgMet 1539  
Dh 4783 ACCGCGCTCTACTCTGCGGGGTGTCCCAACCTGCGCAAGAGCTTCCACATGCAAC 4842  
Qy 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
Dh 4843 CGCGTGTGTGGGTGCAATGCGAACCCTGTAGTGCAGCGCAAAATGTGACATGGCC 4902  
Qy 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLeuAsnValCysAsp 1579  
Dh 4903 GGAATCATGCCAACAATGGACCCGGAAGGCTGCGCTCGGAGAACTTCTGCGAT 4962  
Qy 1580 SerAsnThrCysHisAsnGlyIleThrCysValAlaAsnGlnThrAspAlaPheSerCysGlu 1599  
Dh 4963 GCGAGCGGTGTCAGAATGAGGACCTGTGTCAACAGGTGAAATATGTATGTGTGTAG 5022  
Qy 1600 CysProLeuGlyPheGlyGlyLysSerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
Dh 5023 TGTCACTCCGATTCGCGGGAAGAACTGTAGCAAGACCATGCTCAACCCAGCTCTTC 5082  
Qy 1620 LeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyr 1639  
Dh 5083 AGCGGTGAAGAGGTGTGCTGAGTGAACATCATCATCATCATCTGTGCGCTGTGAC 5142  
Qy 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
Dh 5143 CTGGGCTCATGTTCGGAACCGGAAGAGGACAGGGTCTATATGAGGACCACTGTGT 5202  
Qy 1660 GlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGlyGly 1679  
Dh 5203 GGGCCACCAAGCTTGTGCTCAAGATCTCGAACAACATCTCACTCACTTGAAGTGTCCAC 5262  
Qy 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspThr 1699  
Dh 5263 GGCCTCTCGAATGTGAGTCCGATGATGCTGCGGGTTCGGGTGACGACGCGGAGTGG 5322











Dh	1312	CCGCTCAGTGCACAGCGCCACCGCTGTACATCGAGGTGAGAGACAGAAACGAACTTACCCC	1371
Qy	398	GLInPheSerGIuLVaRgTYrValGLInValArgGIuAaPValThrProGIaValPro	417
Dh	1372	CAGTTTCAGCCGACGAACTACGTGTCTCAAGGTGCCGAGGACGTGGGGCTTCAAACCGGCT	1431
Qy	418	ValLeuAaRgValThrAlaSerAspArgAspIysGIserAaAlaValAlaHisTYrSer	437
Dh	1432	GTGCTTCGAGTGCAGAGCCACGGAACGGAGACAGAGGCCAGAAACGGGCCATTTCATCAAGC	1491
Qy	438	ILewerSerGIyAaAlaArgGIyGLInPheTYrLeuAaPValaGIInThGIyAlaLeuAaP	457
Dh	1492	ATCTCAGCCGGGAAGTGGCCGGCCAGATTCTTACCTGCACCTGCGTAGAGCGGGATCTGGAT	1551
Qy	458	ValValSerProLeuAaAspTYrGIuThrThrIysGIuTYrThrLeuAaRgValArgGLIn	477
Dh	1552	GTGATCAACCCCTTGATTCGAGATTCGAGAGTGTCCAGAAATACCTCGTAGAGATTAAGCCGAG	1611
Qy	478	AspGIyGIaTYrProProLeuSerAaValSerGIyLeuValThrValGLInValLeuAaP	497
Dh	1612	GATGGGGGCCGGCCCCCGCTCATCAATCTTCAAGGGGTGGGTGTGTGTGCAGAGTCTGGAT	1671
Qy	498	ILeasAaPaaAlaProILePheValSerThrProPheGLInaIaThrValLeuGIuSer	517
Dh	1672	GTCAACGACAAACGAGCTTACTTGTGTGACAGCCCTTCAGAGGACCGGTCTCTGAGAAAT	1731
Qy	518	ValProLeuGIyTYrLeuValLeuHisValGLInaIaIleAaPAlaAaPAlaGLIyAaP	537
Dh	1732	GTGCCCTCGGGCTACCCCGTGGTGCACATTCAGCGGTGGACCGGGACTCTGGAGAGAAC	1791
Qy	538	AlaArgLeuGIyTYrArgLeuAla-----GIyValGIyHis--	549
Dh	1792	GCCCCGGCTGACATATCGCTGGTGGACAGCGCTCACTTTCGGGGGGCGGACGCGCT	1851
Qy	550	-----AspPheProPheThrILeasAaRgIyThrGIy	560
Dh	1852	GGCCCTTAAGAACTCTGCCCCCAACCCCTGACTTCCCTTCAGATTCACAAACGCTCGGT	1911
Qy	561	TrpILeSerValaIaValaGIuLeuAaPArgGIuValAaPheTYrSerPheGIyVal	580
Dh	1912	TGGATCAACAGTGTGCGCGAGCTGGACCGCGAGGAGGTGGAGCACTACAGCTTCGGGGGT	1971
Qy	581	GIuAlaArgAaPHisGLIyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal	600
Dh	1972	GAGGGGTGGACCAACGGCTCGCCCCCAAGAGCTCTTCACACAGGTGTTCATCAACGGTGT	2031
Qy	601	LeuAaPValAaAaPAsaAaPProThrPheThrGIInProGIyTYrThrValArgLeuAaP	620
Dh	2032	CTGAGACGTGAATGACACAGACCCGGTGTTCACGACGCCACCTAGACCTTCGTGAAAT	2091
Qy	621	GIuAaPAlaValaValGIyThrSerValaIaThrValSerAlaValaAaPArgAaPAlaHis	640
Dh	2092	GAGGATGGCGGCGTGGGGAGACGCTGTGCTGACCTCGCAGGGCCGGCAGCCGTGAGCCCAAC	2151
Qy	641	SerValILeThrTYrGIInIleThrSerGIyAaThrAaAaArgAaPheSerILeThrSer	660
Dh	2152	AGTGTGATTAACCTACAGCTCAGCGGACGAGCAACCCGGAACCGCTTGTGCATCAGCAAC	2211
Qy	661	GIuSerGIyGIyLeuValSerLeuAlaLeuProLeuAaPTYrLVaLeuGIuArgGLIn	680
Dh	2212	CAGAAAGGGGGCGGCTCATCACTCGCGCTACCTCGACTCAACAAAGAGAGCAG	2271
Qy	681	TYrValLeuAlaValThrAlaSerAspGIyThrArgGLInaPThrAlaGLInIleValaI	700
Dh	2272	TACGTGTGGCGGTACAGATCCAGCGGACAGCGTGTGCACACTCGCAATGTCTTAATC	2331
Qy	701	AaAaValThrAspAlaAaThrHisArgProValPheGIuSerSerHisTYrThrValAaP	720
Dh	2332	AACGTCACTGATGCCAAACCAACCAAGGCGCTGTCTTTCAGAGCTCCCAATTAACAGTGAAT	2391
Qy	721	ValAaRgIuAaPArgProAlaGIyThrThrValValLeuILeSerAlaThrAspGIuAaP	740
Dh	2392	GTCAATGTAGGACAGGCTGTGTGGGCACTTCCATTTGTACCTCAAGGCCAAAGATGAAGAC	2451

QY	741	ThrGluValuSmaIlaAggIIeThrTyPheMeGluAspSerIleProGluPheArgIle	760
Db	2452	AcAGAGAGAGATGCGCGCATCACTACGTGATGAGACCCCGCGCGCATTCGGACTT	2511
QY	761	AspAlaAspThrGlyAlaValAlaThrThcGlnAgluLeuAspTyTyrGluAspGluValSer	780
Db	2512	GAACCCCGACAGTGGCACCATGTATCACCATGATGAGAGCTGAGCTATGAGAACAGTGGCC	2571
QY	781	TyrThrIleuAlaIleThrAlaIleGAspAsnGlyIleProGluIleSerAspThrThrTyr	800
Db	2572	TACACGGCTGACCATCATGAGCCACAGAGACAGCGCATCCCGAGAAATACAGACACACACC	2631
QY	801	LeuGluIleLeuValaAspValaAspAspSmaIaProGluPheLeuArgAspSerTyr	820
Db	2632	CTAAGATCTCTCATCTCCCTCGATGCCAATGACATGACCCAGTTCCTGTGGGATTTCTAC	2691
QY	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGluIleSerAlaThr	840
Db	2692	CAGGGTTCATCTCTGAGAGATGCTCCACCTCGACACAGATCTCCCTCGAGTCTCGCACAG	2751
QY	841	AspArgAspSerGlyLeuAsnGlyArgArgValPheTyrThrPheGlnGlyValAspAspGly	860
Db	2752	GACCGGAGCTCAGGCTCCCATGCGCGCTGTGTACCTTCCAGGGGAGGAGACAGGCG	2811
QY	861	AspGlyAspPheIleValGluSerThrSerGlyIleValAlaArgThrLeuArgArgLeuAsp	880
Db	2812	GATGGGGACTTCTACATCGAGCCACACGTCGGGTGTATGCGACCCAGCGCGGCTGGAC	2871
QY	881	ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspTyGlyMetPro---	899
Db	2872	CGGAGAGATGCGCGGTGACACACTTGGGCTGTGCTGTGATCGGGGAGTCCACT	2931
QY	900	ProAlaArgThrProMetGluValAlaThrValThrValaLeuAspValaAspAspProPro	919
Db	2932	CCCCCTTAGCGCGCTCGGATGAGAAATCCAGGTACATCTTGGACATTATGACAAAGCCCC	2991
QY	920	ValPheGluGluAspGluPheAspValaPheValGluGluAsnSerProIleGlyLeuAla	939
Db	2992	ATGTTTGAGAGAGACGAACCTGAGACTGTGTGTGAGAGAACCAACAGTGGGCTGGGTG	3051
QY	940	ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln	959
Db	3052	GTCGCAAGATTCGTGCTACACAGCCCTGATGAGAGGCCCTTAATGCCAGATCATGTATCAG	3111
QY	960	IleValGluGlyAsnIleProGluValaPheGluLeuAspIlePheSerGlyGluLeuThr	979
Db	3112	ATTGTGAGAGGGAGCATGCGGCACTTCTTCACGTGACCTGCTCAACGGGGAACCTGGCT	3171
QY	980	AlaIleuValaAspLeuAspTyrTyrGluAspArgProGluTyrValaLeuValaIleGlnAlaThr	999
Db	3172	GCCATGTGGAGCTGACCTTGAAGGTCCGGGGAGATGTGCTGTGGTGCAGGACGACG	3231
QY	1000	SerAlaProLeuValSerArgAlaThrValAlaIleValArgLeuLeuAspArgAspAsn	1019
Db	3232	TCCGCTCCGCTGCGTGAAGCCAGACCGATGTGACATCTCTCTCGTGAGACAGAAATGACAC	3291
QY	1020	ProProValaLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValaThrAsnArgSerSer	1039
Db	3292	CCGCTGTGCTGCCGCACTTCACATCTCTTCAACAACATATGTACCAACCAACCTTCACAC	3351
QY	1040	SerPheProGlyValaIleGlyArgArgValProAlaIleAspProAspIleSerAspSer	1059
Db	3352	AGTTTCCCAACCGGCGATGTGCTGTGCATCCCGGCCATATGACCCCGATGTACAGACG	3411
QY	1060	LeuThrTyrSerPheGluArgGlyAsnGluLeuSerIleuValaLeuLeuSmaIaSerThr	1079
Db	3412	CTCAACTACACTTCGAGAGGAGGCAAGAGCTGGCGCTGTGCTGTGAGACCCCGCACAG	3471
QY	1080	GlyGluLeuLeuValaSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSer	1099
Db	3472	GCGCAACTCGGCTCAGCGCGGACCTTGACCAACAGCGCGCTGAGAGCGCTCAATGAG	3531

QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 DB 3532 GGCTCTGTCTGATGATGATCAACAGCTCAACGGCTTCTGCACTCTGGTGTCAACATC 3591  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 DB 3592 ATCAACGAGACATGCTGACCAACACATCACTGCTCCCTGGAGAACATGTCACGAG 3651  
 QY 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
 DB 3652 AAGTTCCTGCTCCCGCTGCTGGCCCTCTGATGAGGGGGTGGCCGCTGGTGTCAAC 3711  
 QY 1160 ProProAspHisValValValPheAsnValGlnAspThrAspAlaProGlyGlyHis 1179  
 DB 3712 ACCAAGAGAGAGCTTCTGCTTCAACGCTCAAGAACACACCACTCC--AGCTTCAAC 3768  
 QY 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyGlyProPhe 1199  
 DB 3769 ATCTGAACTGACCTTCTCGGCGCTGCTGCTGGCGGCTCGCGC-----CACTTC 3822  
 QY 1200 LeuProSerGluAspLeuGlnGlyLeuArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 DB 3823 TTCCTCGTCCGAGGACCTGACGAGGAGCATCTAATCGACGCTGCTGACCAACATC 3882  
 QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
 DB 3883 TTCACGACAGCGGTGCTGCTTCAACGACATCTGCTGCGGACGCTGCGGAC 3942  
 QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
 DB 3943 TACATGAATGCGGTGCTTCTGCAATTCAGACGCTCCGCGCTTCTCTCAAGCTCAC 4002  
 QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
 DB 4003 ACCGAGCTCTTCGCGGCCATCCACCCATCAACGCGCTGCGCTGCGCGCGCGCGC 4062  
 QY 1280 PheThrGlyAspTyrCysGlnThrGlyValAspLeuCysTyrSerArgProCysGlyPro 1299  
 DB 4063 TTCACCGGAGACTACGTCAGACGAGATCGACTCTGCTACTCCGACCCGTCGCGCGC 4122  
 QY 1300 HisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
 DB 4123 AACGCGCGCTGCGCGGCCAGCGCGGCTACCTGCGAGTGTCTTGAGAGACTTCACT 4182  
 QY 1320 GlyGlnHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysGlyAsn 1339  
 DB 4183 GGAGAGCACTGTGAGGTGATGCGCGCTCAGGCGCGCTGCGCAACGCGGTGCGAAGAC 4242  
 QY 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAsp 1359  
 DB 4243 GGGGGGCACTGTGTAACCTGTCACTGCGCGCTTCCACTGCGTGTCTCTCGCGCAG 4302  
 QY 1360 PheGluTyrProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
 DB 4303 TATGAGAGGCTTACTGTAGGTGACACACAGAGCTTCCGCCCAAGTCTTCTGTACC 4362  
 QY 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGlu 1399  
 DB 4363 TTCGGGGGCTGAGACAGCGCTTCCACTTCACTCATCTCCCTCACGTTTGGCATAGAA 4422  
 QY 1400 ArgAspArgLeuLeuLeuTyrThrGlnArgPheAsnGlnTyrHisAlaPheValAlaLeu 1419  
 DB 4423 AGGAACGGCTTCTTCTTCAACGCGCTTCAATGAAACACAGCTTCACTCCCGCTG 4482  
 QY 1420 GluValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
 DB 4483 GAGATGCTGAGAGAGAGGTGACGCTCACTTCTGACGCGGACACACACACACACCGTG 4542  
 QY 1440 SerProPheValProGlyGlyValSerAspGlyGlnThrPheHisValGlnLeuLeuTyr 1459  
 DB 4543 GCACCGAAGGTTCCCAAGTGTGATGACGCGCGGTGCACTGTGACAGGTGACATAC 4602  
 QY 1460 TyrAsnTyrProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnVal 1479

DB 4603 TACAACAAGCCCAATATGCGCACCTGGGCTGCCCCCATGGGCGCTCGGGGAAAGATG 4662  
 QY 1480 AlaValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 DB 4663 GCGGTGTGACAGTGTGATGATGTGACAAACCATGGCTGTGCGCTTGTGAAGACATC 4722  
 QY 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerTyrValSerLeuAspLeu 1519  
 DB 4723 GGGAACTACAGTGTGCTGCCACGAGGACCTCAACCGGCTCCAGAAATCTCTGATCTG 4782  
 QY 1520 ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet 1539  
 DB 4783 ACCGCGCTTCACTCTGCGGGGTGTCCCAACTGCGCAAGAACTTCCAGTGTCAAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnAlaAspSerArgHisIleAspMetAla 1559  
 DB 4843 CGGCAGTGCGGGGTGATGCGGAACCTGTCAAGTCAAGCGCAAAATGTGACATGGCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaTyrGlyAsnValCysAsp 1579  
 DB 4903 GGAATTCATCGCAACAAATGACACCGGAGAGCTGCGCTCGGAGGAACTTCTGCAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGlu 1599  
 DB 4963 GGGAGCGGTGTCAAGATGAGAGCACTGTGTCAACAGTGTGAATGTATCTGTGTAG 5022  
 QY 1600 CysProLeuGlyPheGlyGlyCysSerCysAlaGlnGlyMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTCACTTCCATTCGGGGGGAAGACTGTGAGACGCAAGCCTCAACCCCAAGCTTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyr 1639  
 DB 5083 AGCGGTGAGAGCGTGTGCTTCTGAGTGAAGCTGAACATCATCTGTGTGCTGGTAC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 DB 5143 CTGGGGCTCAATGTCGGAACCGGAGAGAGACAGCGTTCATGAGAGGACCAAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGly 1679  
 DB 5203 GGGGCCACAGCTTCCGCTTCAATCTGAACCAACTCACTCAAGTTGAGTGTCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAlaAsnAspGlyAspTyr 1699  
 DB 5263 GGGCCCTCCAGTGTGAGTCCGTATGCTGCGGGTTCGCGGTGACCGAGGGAGTGG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 DB 5323 CACCACTGCTGATCGAGCTGAAGATGTTAAGAGGACAGTGAATGAACACCTGTGTC 5382  
 QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
 DB 5383 ACCATGACTTGGACTATGGATGAGACCAAGACAGCATATCGGGGGGAGTCTTCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyTyrLeuProGlyProAlaGlyValAla 1755  
 DB 5443 GGGCTGAGAGTAAGAGCGTGTGAGGCGCTCTGAAGACAAGTCTCGGTGCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGluGlyValAsn 1775  
 DB 5503 CGTGAATTCAGAGCTCAAGACAGAGTGAAGATGGGGGGAGCGCCACCAAGTGTGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGluSerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
 DB 5563 ACCGTGAACATGAACAGACGACTCAAGTCAAGGTGAAGACAGCGCTGTGATGTGAGAC 5622  
 QY 1796 ProCysAspSerAspProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815  
 DB 5623 CCTGTACTTCAAGACCTGTGCTCCCAATATAGCGCTGCGCAAGCGCTGGAGAGACTAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835

Db 5683 AGCTGGCTCTGTGACAAAGGCTACCTTGAAATTAAGTGTGTGATGCTGTACCTGAAAC 5742  
 Qy 1836 ProCysgluHhIsgInserValCysThrArgLysProSerAlaProHhIsgLysThrCys 1855  
 Db 5743 CCTGTGGAGAAATGAGGGGCTGTGGCTGTCCCGGCTCCCGGAGGGCTACGTGTGC 5802  
 Qy 1856 GluCysProProbaHhIsgLysThrCysgluThrArgLysLeuPProCysPro 1875  
 Db 5803 GAGTGTGGGCTCCAGTCACTAAGGAGCGGTACTGTGAAGAAACAACTTCACCTTCGTGTCCC 5862  
 Qy 1876 ArgGlyTrpTrpLysHhIsgProThrCysGlyProCysAsnHhIsgLysSerLysGlyPhe 1895  
 Db 5863 AGAGGCTGTGGGGGAAACCCGCTGTGTGACCTGTGCACCTGTCCCTTCAGCAAGGCTTT 5922  
 Qy 1896 AspProAspCysAsnLysThrSerGlyLysCysHhIsgLysGluAsnHhIsgLysPro 1915  
 Db 5923 GATTCGCACTGTATTAAGCAACAGGCGAGTGCACATGCAAGAGAAATTACTCAAGCTC 5982  
 Qy 1916 ProGlySerProThrCysLeuLeuCysAspCysThrProThrGlySerLeuSerArgVal 1935  
 Db 5983 CTAGCCAGGACACTGTGTGCTGTGCTGTGAGCTGTTCCTCCATGAGCTCCACAGCCGACT 6042  
 Qy 1936 CysAspProGluAspGlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAsp 1955  
 Db 6043 TGGGACATGGCCACCGGGCAGTGTGCTGTGACCGGCGCTCATCGGCGCCAGTGCAC 6102  
 Qy 1956 ArgCysAspAsnProPheAlaGluValThrThrAsnGlyCysgluValAsnHhIsgLysSer 1975  
 Db 6103 CGCTGGCAACAACCGTTTGTCCGAGGTCAACAGCTCGGTGTGAAGTATCTCAATGGC 6162  
 Qy 1976 CysProArgAlaIleGluAlaGlyLysThrTrpProArgThrArgPheGlyLeuProAla 1995  
 Db 6163 TGTCCCAAGACATTTGAGCGCGGATCTGTGTGCGACAGACAAAGTTCCGGGACGCGGCT 6222  
 Qy 1996 AlaAlaProCysProLysGlySerPheGlyThrAlaValArgHhIsgCysAspGluHhIsg 2015  
 Db 6223 GCGGTGCAAGCTCTTAAGGATCCGTTGGAATGCGGTCCGACACTCCAGCGGGAGAAAG 6282  
 Qy 2016 GlyTrpLeuProProbaHhIsgLysThrSerLysThrPheSerGluLeuLysGly 2035  
 Db 6283 GGGTGTGTGCTCCAGAGCTCTTAACTGTATCAACCAATCTTCGTCGAGCTCAAGGGCC 6342  
 Qy 2036 PheAlaGluArgLeuGlnArgAsnGlnSerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 Db 6343 ATGAATGAGAGGCTGAGCCGCAATGAGACGAGGTGACGCGCGCCAGGCGCTGTGACGCTG 6402  
 Qy 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHhIsgThrAlaGlyTrpPheGlySerAspValLys 2075  
 Db 6403 GTGAGGGCGCTGGGCAGTGTATACAGCACACGGGCAAGCTTTGGCAATGACGTGCGC 6462  
 Qy 2076 ValAlaTrpGlnLeuAlaThrArgLeuLeuAlaHhIsgLysThrGlnArgLysGly 2095  
 Db 6463 ACGGCTTACAGAGCTGTGGGCAAGCTCTTACAGCACAGAGCTGTGAGCGGCTTGCAC 6522  
 Qy 2096 LeuSerAlaThrGlnAspValHhIsgPheThrGluAsnLeuLysGlyValIleGlySerAlaLeu 2115  
 Db 6523 CTGGCAAGCAAGGAGGAGCGGCACTTCAAGAGAGCTCATCACTCGGCGAGCGGCTC 6582  
 Qy 2116 LeuAspThrAlaAsnLysArgHhIsgTrpGluLeuIleGlnGlnThrGlnGluGlyThrAla 2135  
 Db 6583 CTGGCCCCACGACACAGGGGGGCTGGGAGAGATCCAGCGGAGCGAGGGCGGCAAGGCA 6642  
 Qy 2136 TrpLeuLeuGlnHhIsgLysGluAlaTrpAlaSerAlaLeuAlaGlnAsnMetArgHhIsg 2155  
 Db 6643 CAGTGTCTCCGGCTCTGAGGGCTACTTACAGCAACGTGGCAAGCAAGTGGCGGAGAG 6702  
 Qy 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
 Db 6703 TACCTGGGGCTTCTGTGATGTACCGGCAACATGATCTGTGTGTGACATCTTGAC 6762  
 Qy 2176 LysGlyAsnPheAlaGlyAlaLysLeuProArgTrpGlyLysAlaLeuArgLysGlnPro 2195  
 Db 6763 AAGTTCAACTTACGGGAGCGAGGCTCCGGATTCGACACACATCATGAAGATTTCCC 6822

Qy 2196 ProAspLeuGluThrThrValIleLeuProGlnSerValPheArg-----Glu 2211  
 Db 6823 AGGAGAGCTGGAGTCTCTCCGCTCTCTCCAGCGGCACTTCTTACAGACCACTGAAGAAAA 6882  
 Qy 2212 ThrProProValAlaArgProAlaGly-----ProGly 2222  
 Db 6883 GAAAGCCCCCTGTGAGGCGGCTGTGGCGGAGAACACCCCGGACCAAGCGGCTGGGG 6942  
 Qy 2223 GluAlaGlnGluProGluGluLeuAlaArgArgGlnArgHhIsgProGluLeuSerGln 2242  
 Db 6943 CCGGACACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7002  
 Qy 2243 GlyGluAlaValAlaSerValIleIleTrpArgThrLeuAlaGlyLeuLeuProHhIsgAsn 2262  
 Db 7003 CAGTTTGGCGGTGCTGTGTATCACTTAACGCAACCTGGGGGAGAGCTCTGCGCCAGAGCGC 7062  
 Qy 2263 TyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThrProVal 2282  
 Db 7063 TACGACCCCGAGCGGTGAGCGCTCGGTGTGCTCAACCGGCCCATATTAATACCCGATG 7122  
 Qy 2283 ValSerIleSerValHhIsgAspAspGluLeuLeuLeuProArgAlaLeuAspLysProVal 2302  
 Db 7123 GTGACACAGCTGTGTATACAGGAGGGGCTCCGCTCCGAGAACCTCTGAGAGGCGCTTC 7182  
 Qy 2303 ThrValGlnPheArgLeuLeuGlnThrGluArgThrLysProIleCysValPheTrp 2322  
 Db 7183 CTGTGTGAGTTTGCCTGTGTGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7242  
 Qy 2322 AsnHhIsgSerIleLeuValSerGlyThrGlyTrpSerAlaArgGlyCysgluValVal 2342  
 Db 7243 AACCACTCCCTGGCCGTTGTGTGAGAGGAGGTGTGTGCTCCGAGGCTGTGCGACTCTCG 7302  
 Qy 2343 PheArgAsnGluSerHhIsgValSerCysGlnCysAsnHhIsgMetThrSerPheAlaValLeu 2362  
 Db 7303 TCCAGGAAACCGGACACATGTGCTGTGACAGTGCACCAACAGCGAGCTTTCGGTGTGC 7362  
 Qy 2363 MetAspValSerArgArgGluAsnGlyLysLeuPProLeuLysThrLeuThrVal 2382  
 Db 7363 ATGATATCTCCAGCGCGTGAACCGGAGGTCTCGCTGTGAATGTCTACATATGCC 7422  
 Qy 2383 AlaLeuGlyValThrLeuAlaIleLeuLeuLeuThrPhePheLeuThrLeuLeuArg 2402  
 Db 7423 GCTGTGTCTTCTACCTGTGACAGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7482  
 Qy 2403 IleLeuArgSerAsnGlnHhIsgLysLeaArgAsnLeuThrAlaIleLeuLysLeuAla 2422  
 Db 7483 AGCTGGCTCCACCTGACAGATTCACAGACCTGCGGTGGCGCTTCTCTCTCT 7542  
 Qy 2423 GlnLeuValPheLeuLeuGlyLysGlnGlnAlaAspLeuProPheAlaCysThrValIle 2442  
 Db 7543 CAGCTGTGTGTGTATTTGGATTCACAGAACCGGTTCTGTGCAAGTGTGT 7602  
 Qy 2443 AlaIleLeuLeuHhIsgPheLeuTrpLysCysThrPheSerTrpAlaLeuLeuGluAlaLeu 2462  
 Db 7603 GCGATCTCTCTCACTTACTATCTACATGACACCTTGTCTGTGACCTGTGTGAGAGCTGTG 7662  
 Qy 2463 HhIsgLysArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTrp 2482  
 Db 7663 CATGTCTACCGGATCTGACGAGGTGCGCAACATGCAACGAGGCGCATGTGGGTTCTTAC 7722  
 Qy 2483 TyrMetLeuGlyTrpGlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspPro 2502  
 Db 7723 TACGT 7782  
 Qy 2503 GlyGlyTrpGlyLysProAspPheCysTrpLeuSerIleLysAspThrLeuIleTrpSer 2522  
 Db 7783 CAGGCTTACGAGAACCCGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7842  
 Qy 2523 PheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrpIleLeuAlaAla 2542  
 Db 7843 TTTGGGGGGCCCATGAGAGCTGTATTAATCATCAACAGATGCACTTGTGTCTTATCTGCA 7902













OY	2608	SerTyrValIleuSerArgGluValAlaArgValAlaLeuLys-----LeuAlaCysSerArg	2656
Db	8143	TTCACATCGCTGTCTCCAACCAAGAGCTCCGAAGAACAATTAAAGGGCGTGCTCGCGGAGAG	8202
OY	2627	LysProSerProPheProAlaLeuThrThirLysSerThr---LeuthrSerSerTyrAsn	2645
Db	8203	AAGCTGCACCTGAGAGACTCCGCCAACCACAGGGCACCTGTGACCGGCTCTCCAC	8262
OY	2646	CysProSerProTyrAlaAlaArgGly-----ArgLeuTyrGlnProTyrGlyAspSerAla	2663
Db	8263	TGCAAACACCACTTCGTGGTAGCGGGCTGACANGTCTGCGCACAGACTTTGGCGAGTCCACC	8322
OY	2664	GlySerLeuHisSerThrSerArgSer-----	2672
Db	8323	GCCTCGCTGCACAGCATGTCTACGGGATGAAGGATCCAGAAAGCTCGGCTCTCTGG	8382
OY	2673	-----GlyLysSerGlnProSerTyrTrileProPheLeuAlaArgGlu	2686
Db	8383	CTGTGTAGAGGGACAGCCAGAGAGCCAGACCGCTCCCTCAAGCCCAAGAGCTGCAGAAAT	8442
OY	2687	GluSerAlaLeuAnProGlyGlnGlyProProGlyLeuGlyAspProGlySerLeuPhe	2706
Db	8443	-----CCCCCTGGC-----	8451
OY	2707	LeuGlnGlyAlaArgGlnGlnHisAspProAspThrAspSerAspSerLeu	2726
Db	8452	-----CACCATTCGCACTCAGATTAAGGAACTGTCTCC	8484
OY	2727	GluAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGlnGlnGln	2746
Db	8485	--GATGAGCAAGACACTTACGCTTCCTCCACACTGTGTACAGAGAGAAAGATAGG	8541
OY	2747	GlnGlnGlnGlnGlnGlnGlnAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGly	2766
Db	8542	GTGGAGACTGAGAAAA-----TGGAGCCCGCGCAAGGGCC	8577
OY	2767	ProGlyAlaGluArgLeuProLeuHisSerThrProLys---AspGlyGlyProGlyPro	2785
Db	8578	-----GCCGTCCACAGCAACCCCAAAGGAGACGCTGTGSCAACAC	8619
OY	2786	GlyLysAlaProTyrProGlyAspPheGlyThrThrAlaLysGlnSerSerGlyValAnGly	2805
Db	8620	GTTCCGGCGCGCTGGCCCGACCAAGACGCTGAGAGTACATAGAGAACCCACGGCC	8679
OY	2806	AlaPro-----GlnGluArgLeu	2811
Db	8680	AAGCCCGCCCTGAAGGTGAGACCAAGGTCAAGCTGAGACTGCACCGGAGAGAACAGGGC	8739
OY	2812	ArgGluAnGlyAlaPheAlaLeuSerArgGlnGlySerLeuGlyProLeuProGlySerSer	2831
Db	8740	AGTCAACCTGTAGAGATTACCCCCCGGACCAAGAGAACGGGGGCGCAGCCAGGCTGTACGC	8799
OY	2832	AlaGlnPro-----HisLysGlyTlleuLysValLysCysLeu-----Pro	2845
Db	8800	AGCAGACCCCCAGAGACAGAGAGGAGCATTTGAAAATAATAAATGACACTACCCGCGCGCTG	8859
OY	2846	ThrIleSerGlnLysSer-----SerLeuLeuArgLeuProLeuGlnGlnCysThrGly	2863
Db	8860	ACGCTGACGAGACACACTGAAGGGCGCGCTCCCGGAGAAAGCTGGCCAGCTGTAGACAG	8919
OY	2864	SerSerArgGlySerSerAlaSerGlnGlySerArgGlyGlyPro-----Pro	2879
Db	8920	AGCCCAATCTCTCGCGACAGCTTCCCTCGGGCTCTGGCGGCCCGACCTGGCCATCACA	8979
OY	2880	ProArgProProProArgGlnSerLeuGlnGlnGlnLeuAnGlyValMetProIleAla	2899
Db	8980	GTCAGAGACCCCTGGAGAGAGCGGGGCGTACACCTTCAACGGGGT-----GCC	9030
OY	2900	MetSerIleLysAlaGlyThrValaArgGluAspSerSerGlySerGln	2915
Db	9031	ATGAAATGTGGCACTGGAGAGCGCCAGGCCAATGACTCGACTCTGAG	9078

ID	AAT85320		standard; cDNA; 10195 BP.
XX	AAT85320:		
AC			
DQ			
DT	08-DEC-1997	(first entry)	
XX			
DE	Mouse receptor ME2 encoding cDNA.		
XX			
KW	Chemical screening; epidermal growth factor; BGF, repeat; cadherin;		
KM	neuronal cell division; differentiation; cell physiology; trauma;		
RW	neurodegenerative disease; adenylate cyclase; CAMP; ss.		
XX			
OS	Mus masculinus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	2..8125	
F7		/tag= a	
F7		/product= ME2	
PX	MO9707209-AZ.		
PN			
PD	27-FEB-1997.		
PE	14-AUG-1996;	96WO-GB01980.	
PR	21-AUG-1995;	95GB-0017098.	
PA	(UNLO ) IMPERIAL COLLEGE SCT TECHNOLOGY & MED.		
PL	Hadjantonakis A, Little P,		
DR	WPI; 1997-165293/15.		
PS	P-PADB; AA#27161.		
PT	Screening chemicals which interact with receptor contg. EGF-like		
FT	and/or cadherin-like repeats - to develop pharmaceuticals for		
XX	treating neurodegeneration, trauma, etc.		
PS	Claim 12; Fig 9; 55pp; English.		
XX			
CC	A method has been developed for the screening of a chemical for		
CC	subsequent use as a pharmaceutical agent. The method involves contacting		
CC	the chemical with a receptor comprising epidermal growth factor		
CC	(EGF)-like repeats and/or cadherin-like repeats and determining whether		
CC	the chemical interacts with the receptor to form a complex. The present		
CC	sequence encodes the mouse receptor ME2, which can be used in the above		
CC	method. The method is used to screen for chemicals which can cause		
CC	neural cells to divide or differentiate, can affect cellular physiology,		
CC	repeat trauma, treat neurodegenerative disease, stimulate adenylyte		
CC	cyclase production, increase cAMP levels and promote neural growth.		
SQ	Sequence 10195 BP; 2265 A; 2981 C; 2855 G; 2094 T; 0 other;		
XX			
XX			
Alignment Scores:			
Pred. No.:	0	Length:	10195
Scores:	8710.00	Matches:	1668
Percent Similarity:	74.41%	Conservative:	408
Best Local Similarity:	59.78%	Mismatches:	601
Query Match:	56.03%	Indels:	116
DB:	18	Gaps:	29
US-09-916-849A-3	(1-2923) x AAT85320	(1-10195)	
OY	192 GLUASGNLPROALLAGLYTHRProValAlaserleuArgAlalleappProAspgIngly	2111	
Db	104 GAACGGAACCGTGGCGCACTGCCGTTCATCNAACTGCAGCGCACGACCAGACGAAAGC	163	
OY	212 GLUALGLVARGLeuglutyThrMetapalaleupehapserarvgsSerasgnPhe	2311	
Db	164 GATCGAGAGCCCTTAGCTAACCATMGAGAGCGCTTTCCATTGAGCGCTTAATGSGCTAC	2233	
OY	232 PheserleuaspprovalThrclValalaValThrThralaglugluueasparvgIutThr	2511	

Db 224 TTCCTCATGATGCGCCGACGAGGTGACGACCGCCCTCCCTGGACCGGAGAAC 283  
Qy 252 LysSerThrHisValPheArgValThrAlaGlnAspHisGlyMetProArgArgSerAla 271  
Db 284 AAGACACTCACTGATCTCAAGTATGCTGTGGACCAAGCTCCCGAGGGCTCAAGCT 343  
Qy 272 LeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAspProValPheGlnGln 291  
Db 344 GCCACCTACCTCACTGATGACCTAAGCACTCAAGCAAGCCCACTGCTTGAAGCAG 403  
Qy 292 GlnGluTyrLeuGlnSerLeuArgGluAsnLeuGlnValGlyTyrGlnValLeuThrVal 311  
Db 404 TCTGATATCGAGAGGAACTCGAAGAAACCTGAGAGTGGCTATATAGGTCTTGACATC 463  
Qy 312 ArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyrArgLeuGlnGly 331  
Db 464 CGTGCACGACGAGGGATGCCCTTCCACGCAACATGCGCTATGCTGTGAGAGGC 523  
Qy 332 SerGlyGlySerProSerGlnValPheGluIleAspProArgSerGlyValIleArgThr 351  
Db 524 GCAAGGTGT-----GTCTTGGATAGACGACGATCAGGTGCGTGCGCAC 571  
Qy 352 ArgGlyProValAspArgGlnGlnValGluSerTyrGlnIleuThrValGlnIleSerAsp 371  
Db 572 CGAGCTGTGTGACCGTGAAGAGCGGCTGAGTACAGCTGCTGTGAGGCGCAATAGC 631  
Qy 372 GlnGlyArgAspProGlyProArgSerThrThrAlaAlaValPheLeuSerValGlnAsp 391  
Db 632 CAGGCTCGGAATCGAGGCCCACTAGTGTCTAGCACCCTCAACATAGTGTATGAAGAC 691  
Qy 392 AspAsnAspAsnAlaProGlnPheSerGlyValArgTyrValIleGlnValArgGlnAsp 411  
Db 692 GAGAAATGACACTACCCCGAGTTCAGTGAAGACCCCTATGTGTTCAGTCCCAAGAAC 751  
Qy 412 ValThrProGlyAlaProValLeuArgValThrAlaSerAspAspAspGlySerAsn 431  
Db 752 GTAGCCGTCAACACGCGTGTGCTGAGTCCAGGCCATGACCGGACCGAGGCGCAAT 811  
Qy 432 AlaValValHisTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAla 451  
Db 812 GCAAGCATACACTACGATCGTATGTGCAACCTGAAGGGTCAAGTTCACCTGCAATTCG 871  
Qy 452 GlnThrGlyAlaLeuAspValIleSerProLeuAspTyrGlnThrThrIleGlyIleTyrThr 471  
Db 872 CTTAATGGAGCTGATGATATCAACCGCTGACCTTCGAAGCCATCCGGGAAATACACC 931  
Qy 472 LeuArgValArgAlaGlnAspGlyArgProProLeuSerAsnValSerGlyLeuVal 491  
Db 932 CTGCGCATCAAAAGCCCAAGATGGGGCGGCGCTCTCTCATTAATTCCTCAAGACTGTGC 991  
Qy 492 ThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGln 511  
Db 992 TCGGTGACAGGTGTATGATGTAACACATAGCGCCATCTTGTATGACAGCCCTTTCAG 1051  
Qy 512 AlaThrValLeuGlnSerValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAsp 531  
Db 1052 GGTGCGGTCTGAGAAATGTGCCCTCGGCACTGACATCTCTGACATCCAAAGCGGTGAGC 1111  
Qy 532 AlaAspAlaGlyAspAsnAlaArgLeuGlnTyrArgLeuAlaGlyValGlyHis----- 549  
Db 1112 GCAAGTACAGGGAGAACCCAGGCTGACATCCGCTATGTGACACAGCTCCCATATC 1171  
Qy 550 -----AspPheProPheThr 554  
Db 1172 GTGGGGGAGCAGTGTGACTCTGAGAACCTGTGCTGCGCCAGACTTCCCTTCCAA 1231  
Qy 555 IleAsnAspGlyThrGlyTyrIleSerValAlaAlaGlnLeuAspArgGlnValAsp 574  
Db 1232 ATCCCAACAACGCTCGGTGATTACTGTGTGCGGAGCTGGAACCGTGAAGAGGTGAA 1291  
Qy 575 PheTyrSerPheGlyValGlnValIleArgAspHisGlyTyrProAlaLeuThrAlaSerAla 594

Db 1292 CACTATACCTTTGAGATGAAGACGTGGAACAATGCTTCAACCAAGCATAGCTCTGTGCC 1351  
Qy 595 SerValSerValThrValLeuAspValAsnAspAsnProThrPheThrGlnProGlu 614  
Db 1352 AGCGTGTCACTACAGTCTGTGATGTATATATTAACAGACCCCATTTACAGAGCTGTG 1411  
Qy 615 TyrThrValArgLeuAsnGlnAspAlaAlaValGlyThrSerValValThrValSerAla 634  
Db 1412 TATGAGCTGCGTGAATAGAGATGCGGCTGTGCGGAGACAGAGTCTGACCTCAGAGGCC 1471  
Qy 635 ValAspAspAspAlaHisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsn 654  
Db 1472 CGAGACCGTGAATGCCAATATGTATATCACTTACCAAGCTGACGCGGTGGAACAACCGGCAC 1531  
Qy 655 ArgPheSerIleThrSerGlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAsp 674  
Db 1532 CGCTTGCACATCAGAGCCAGACAGCGCGGTGCTTATCACCCTTGACCTGCGCTGAGC 1591  
Qy 675 TyrIleLeuGlnArgGlnTyrValLeuAlaValThrAlaSerAspGlyTyrArgGlnAsp 694  
Db 1592 TACAAGCAGGAACGCGATATGTGCTGTGACCGGCTCGATGGCAGCGCTTCAAC 1651  
Qy 695 ThrAlaGlnIleValAlaAsnValThrAspAlaAsnThrHisArgProValPheGlnSer 714  
Db 1652 ACCGCGAGTCTTATATCAAGATTACAGATGCCAACCAACGACGCGGTTTCCAGAGT 1711  
Qy 715 SerHisTyrThrValAsnValAsnValAsnAspArgProAlaGlyTyrThrValValLeuIle 734  
Db 1712 TCCCACTACAGCTCAAGTGTATGAAGACCGGCGCGGTGACCTCACTGATGATCAATC 1771  
Qy 735 SerAlaThrAspGluAspThrGlyGlnAsnAlaArgIleThrTyrPheMetGluAspSer 754  
Db 1772 AGTGCACGATGAGATACCGGTGAGAACCGCGCATCACTATGTGCTGAGAGATCCC 1831  
Qy 755 IleProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeuAsp 774  
Db 1832 GTACACAGTTCGGATGATGACCGGACCTGGACCATTTATACATGACGAAATGTGAC 1891  
Qy 775 TyrGluAspGlnValSerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGln 794  
Db 1892 TATGAGACCAAGCGTGCCTTACAGCTGGCCATCAAGCTGACGAAATGGATTCCTCAG 1951  
Qy 795 LysSerAspThrThrTyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGln 814  
Db 1952 AAGTCAACACTACCTCTCTGAGATCTTATCTCGACCGCAATGACCAAGCGCGCCAG 2011  
Qy 815 PheLeuArgAspSerTyrGlnGlySerValTyrGluAspValProProPheThrSerVal 834  
Db 2012 TTCCTGCAATTTCTACCAAGGATTCGTGTTTCAGAGATGCCCCCACTCAACAGTGTG 2071  
Qy 835 LeuGlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPhe 854  
Db 2072 CTCACAGTCTGTGTTACAGACCGTACAGCGCCCTATATGCGCGCTCTGTACACTTTC 2131  
Qy 855 GlnGlyGlyAspAspGlyAspGlyAspPheIleValGlnSerThrSerGlyIleValArg 874  
Db 2132 CAGGTGGGATGATGAGATGAGATTTCTACATTTGAGCCCAAGTCTGTGTATCCGT 2191  
Qy 875 ThrLeuArgArgLeuAspArgGluAsnValAlaGlnTyrValLeuAspAlaTyrAlaVal 894  
Db 2192 ACCCAAGCCCGGCTGACAGAGAAATGTGCCCTGTGACCAACTTTGGGCTCTGCGTGTG 2251  
Qy 895 AspIleGlyMetPro-----ProAlaArgThrProMetGlnValThrValThrValLeuAsp 913  
Db 2252 GATCGGGGAGCCCGAATCCCTCAGTGCCTCAATGGGAATTCAGTGTAGTGTGTGAC 2311  
Qy 914 ValAsnAspAsnProProValPheGlnGlnAspGluPheAspValPheValGlnGlnAsn 933  
Db 2312 ATTTACCAACAACCCCCAGGTGTTGAGAAAGCAAGCTGAGCTGTGTTGAGAAAGAAC 2371  
Qy 934 SerProIleGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyTyrAsn 953  
Db 2372 AGCCCTGTGGGTTCAGTGTGACAAATTAAGGCCCAACAGACCGGACGAAGTCCGAAAT 2431

QY 954 AAGGlnIleMetYrGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIle 973  
 Db 2432 GCTCAATGATTTATTCAGATCGTGAAGGCAATGTCCGAGGCTTCGAGCTGGACCTA 2491  
 QY 974 PheSerGlyLeuLeuThrAlaLeuValAspLeuAspYrGluLeuAspArgProGluYrVal 993  
 Db 2492 CTGAGTGTGATCCCTCGCTGCTGTGAGTGTGATTTGAGTGTCCGAGGAGCTAATAT 2551  
 QY 994 LeuValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1013  
 Db 2552 TTGGTGTGAGGCACTGTCTCTCTGTGTAAGCCGGCCACCGTGCACATCCGCTCTC 2611  
 QY 1014 LeuAspArgAsnAspAsnProProValLeuGlyAsnAspGluIleLeuPheAsnAspYr 1033  
 Db 2612 CTGACACAGATGCAACCAACCGAGATGTGCTACTTCCAGATCTTTTCAACCACTAT 2671  
 QY 1034 ValThrAsnArgSerSerSerPheProGlyAlaAlaIleGlyArgValProAlaHisAsp 1053  
 Db 2672 GTCAACCAATATTCACAGCTTCCCAAGTGTGTATGAGCGCCATCCAGCCCAAGAC 2731  
 QY 1054 ProAspIleSerAspSerLeuThrYrSerPheGluArgGlyAsnGluLeuSerLeuVal 1073  
 Db 2732 CTGACCTATCTGACACCTCTAATTAACCTTTCTGCAAGGCAACGAGCTGAGCTGCTG 2791  
 QY 1074 LeuLeuAsnAlaSerThrGlyGlyLeuLeuLeuSerArgAlaLeuAspAsnAspArgPro 1093  
 Db 2792 CTGTGTGATCCCGCCAGAGAGATTCAGCTCAGCCGGAGTCTGCAACAACCCGGCCA 2851  
 QY 1094 LeuGlnAlaIleMetSerValLeuValSerAspGlyValHisSerValThrAlaGlnCys 1113  
 Db 2852 CTGAGGCGCTCAGAGAGTGTCTGTCTCAGATGTATCCACAGCCGCTCCTCTGTC 2911  
 QY 1114 AlaLeuArgValThrIleLeuAspGluMetLeuThrHisSerIleThrLeuArgLeu 1133  
 Db 2912 ACTGTGCGCTGACCATTAAGATGACATGTGTACCAAGCAACATCTGCTCCGCTG 2971  
 QY 1134 GluAspMetSerProGluLeuPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaVal 1153  
 Db 2972 GAGAAACATGTGCGAGAGAGATCTGTCCCGCTGCTCCCTTTGAGAAAGGGGTG 3031  
 QY 1154 AlaAlaThrLeuAlaThrProProAspHisValAlaValPheAsnValGlnAspArgThr 1173  
 Db 3032 GCCACAGTACTGTCCACCAAGAGATGACATCTTCGCTTCAACATCCAGAACAGACG 3091  
 QY 1174 AspAlaProGlyGlyHisIleLeuAsnValSerLeuSerValGlyGlnProProGlyPro 1193  
 Db 3092 GAGCTC---AGCTCCACATCTGACATGATCTTCTCGGACATGCTCCCGGTGGACCC 3148  
 QY 1194 GlyGlyGlyProProPheLeuProSerGluAspLeuGlnGlyLeuArgLeuYrLeuAsnArg 1213  
 Db 3149 CGTGGC-----CGGTCTTCCCGTCAGAGGACCTGAGGAGCACTCAACCGG 3202  
 QY 1214 SerLeuLeuThrAlaIleSerAlaGlyArgValLeuProPheAspAspAsnIleCysLeu 1233  
 Db 3203 ACATCTGCTCACCACTTCGCCCAAGGTGTGCTGCTTGTATGACACATCTGCTG 3262  
 QY 1234 ArgGluProCysGluAsnYrMetArgCysValSerValLeuAspPheAspSerAla 1253  
 Db 3263 AGGAGAGCCCTGCGAAGTCAATGAGTGTGCTCGGTGCTTGTGACAGTTCCGCA 3322  
 QY 1254 ProPheIleAlaSerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArg 1273  
 Db 3323 CCTTCCTTATGATTCACACAGTGTCTCTTCGAGCTTCCATCCATCAACGAGCTTCGCG 3382  
 QY 1274 CysArgCysProProGlyPheThrGlyAspYrCysGluThrGluValAspLeuCysYr 1293  
 Db 3383 TGCCGCTGCGCGCGGTTTACCGGGAGCTACTGCGAGACTGAAATGATCTTGTCTAC 3442  
 QY 1294 SerArgProCysGlyProHisGlyArgCysArgSerArgGluGlyYrYrThrCysLeu 1313  
 Db 3443 TCCATTCGCTGCGGGGCAATGCGGCTGCGGAGCCGTGAGGGTGGCTACACTTGTGAG 3502

QY 1314 CysArgAspGlyYrThrGlyGluHisCysGluValSerAlaArgSerGlyYrArgCysYrThr 1333  
 Db 3503 TGCTTGAGAGACTTCACTGTGGGAGCATTCGCCGATGAACTTCCGCTCAGGCGCGCTGTGCC 3562  
 QY 1334 ProGlyValCysValAsnGlyGlyYrThrCysValAsnLeuLeuValGlyGlyPheCysCys 1353  
 Db 3563 AGCGAGTGTGCAAAACCGGGGCACTGTGCAACCTGTCTCAATTTGAGGCTTCCACTGT 3622  
 QY 1354 AspCysProSerGlyAspPheGluYrProYrCysGlnValThrArgSerPhePro 1373  
 Db 3623 GGTGTCCCGCCCGCGAGATATGACATCCCTTACTGTGAATGACACAGACCTTCCCA 3682  
 QY 1374 AlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeu 1393  
 Db 3683 CCCAGTCTCTGTATCTTCCGAGGCTGCGGCAACGCTTCACTTCACTCCCTCTCCCTG 3742  
 QY 1394 SerPheAlaThrGlyGluArgAspGlyLeuLeuLeuYrAsnGlyYrAspPheAsnGlyCys 1413  
 Db 3743 GCGTTGCAACCGAGGACAGAAATGCGCTGTCTCTTACATGAGCCGCTTCAATGAGAG 3802  
 QY 1414 HisAspPheValAlaLeuGluValIleGlnGlnValAlaGlnLeuThrPheSerAlaGly 1433  
 Db 3803 CAGACTTATGATGCCCTGAGATTTGTGAGAGAGAGCTGCACTCACTTCTCGGAGGT 3862  
 QY 1434 GluSerThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrHis 1453  
 Db 3863 GAGACCAACAACCGGTGACACCGCAGCTTCTGAGGTGTGAGCGATGGCGGTGGCAT 3922  
 QY 1454 ThrValGlnLeuYrYrThrAsnYrProLeuLeuGlyGlnThrGlyLeuProGlnGly 1473  
 Db 3923 TCGGTCTGTGTGAGTACTTACACAAAGCCCAATTTGGCCACTGTGGGCTTCCCGCCAGG 3982  
 QY 1474 ProSerGluGlnLeuValAlaValAlaValThrValAspGlyCysAspThrGlyValAlaLeu 1493  
 Db 3983 CGGTGTGAGAGAGAGT 4042  
 QY 1494 ArgPheGlySerValLeuGlyAsnYrYrSerCysAlaAlaGlnGlyYrThrGlnGlySer 1513  
 Db 4043 CACTTTGAGAGTTCCTGTGGGAACTACACATGCGCTCCAGGCACTCAAGACGGCTCC 4102  
 QY 1514 LysLysSerLeuAspLeuThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGlu 1533  
 Db 4103 AAGAACTTACTGTGATGTGCTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4162  
 QY 1534 SerPheProValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSer 1553  
 Db 4163 GACTTCCCGTGACAGCGCTCAGTTGTGGAGTGCATGCAAACTGTTCATCGATGGC 4222  
 QY 1554 ArgHisIleAspMetAlaAspPheIleAlaAsnAsnGlyThrValProGlyCysProAla 1573  
 Db 4223 CGGATTTGTGACATGGCTCGTTTATTTGCCAACAATGTATCCAGGGCAGGCTGTGCTTCT 4282  
 QY 1574 LysLysAsnValCysAspSerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrp 1593  
 Db 4283 CAGAGAACTTCTGCGATGTGACCTTATGCGAGAGCGGGGCACTGTGTGAACAGCTGG 4342  
 QY 1594 AspAlaPheSerCysGluCysProLeuGlyPheGlyGlyYrSerCysAlaGlnGlnMet 1613  
 Db 4343 AACACGTAATAATGATGAGTGGCGCTCCGCTTGTGGGAAAGAACTGTGAACAAGCTATG 4402  
 QY 1614 AlaAsnProGlnHisPheLeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuPro 1633  
 Db 4403 CCAACCCCTTAGGCTTCACTGTGAGACGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4462  
 QY 1634 IleSerGlnProTrpYrLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeu 1653  
 Db 4463 ATCTCTGTGCTTGTGATCTGTGGGCTCATGTTCCGAGCCCGAAGAGATGTGTGTGTGTGT 4522  
 QY 1654 LeuGlnAlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisVal 1673  
 Db 4523 ATGAAAGCAACAGCTGCAAGTCTTCCAGGCTCCATCTCCAGATTTCTCAACAGCTCAATC 4582  
 QY 1674 MetLeuSerValGluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArg 1693



```

Db      4583  CCCTTTAGAGTCTCTACGCGCCCTCTGACGTGGCACTGCATGCAGCTGTCCAACTCCCG 4642
Qy      1694  ALaAenAPGlyVaerTriphishialaLeuGlyAlaSerGlyLeuProGly 1713
Db      4643  ALaAenAPGlyVaerTriphishialaLeuGlyAlaSerGlyLeuProGly 1713
Qy      1714  Hs-----AlaLeuSerPheAPGlyGlyGlnGlnArgAlaGly 1728
Db      4700  AAGACATCAATATACCTGCACTGATCATGCTTGAATGAGTGAACAGACAGT 4759
Qy      1729  AenLeuGlyProArgLeuGlyLeuGlyLeuSerAenLeuGlyValGlyLeuPro 1748
Db      4760  CAGATTGGGAATCAAGCTTCTGGGTGAAGATGGCACTATTGTCACTCGAGGTGGAC 4819
Qy      1749  GlyProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSer 1768
Db      4820  GAGGACAAAGTCTCTCTCCGCACTGTTCCAGAGCTGTATGCAAGGAGTGGAGTGG 4879
Qy      1769  AepTriProGlnGlyValAenSerLeuAPProSerHisGlyLeuSerIleAenValGlu 1788
Db      4880  GAGAGCTCCACCAACATGCTCCACCTGAACTGAATGACGCTCAAGGTCAAGGTGAAG 4939
Qy      1789  GlnGlyCysSerLeuProAPProCysAenSerAenProCysProAlaAenSerTyCys 1808
Db      4940  GACGCTGTGATGTGAGAGCAACATGTGCTCAAGCCCTGCTCCCATGACCCCTGC 4999
Qy      1809  SerAenAPTriAPSerTySerCysSerCysAenProGlyTyTyGlyValAenCys 1828
Db      5000  CGTGACACATGGGACAGCTACTCTGCACTGTGACAGAGGTACTGGAAAAAGTGT 5059
Qy      1829  ThrAenValCysAenLeuAenProCysGlnHisGlnSerValCysThrArgTyProSer 1848
Db      5060  GTGAGTCACTGTCTCTGTAACCTCTGCAAGCAAGTGGACAGCTGTGTGCTCCCAAC 5119
Qy      1849  AlaProHisGlyTyTyThrCysGlnCysProProAenTyLeuGlyProTyCysGlyThr 1868
Db      5120  ACTCTCGAGGCTACTCTGCAAGTGTGAGCCGCGCACTATGAGCAAGTGTGAGAAC 5179
Qy      1869  ArgIleAPGlnProCysAenProArgGlyTyTyGlyValProThrCysGlyProCysAen 1888
Db      5180  AAGGTGACCTTCCGTGCGCCCAAGGCTGTGTGGGAA-CCGTGTGTGCGCCCTGTAC 5238
Qy      1889  CysAenValSerTyLeuPheAenProAPCysAenTyTySerGlyGlyCysHisCys 1908
Db      5239  TGTGTGTGAGCCAGGCTTGTATCCGCACTCAACAAAGCAAGCCAG-----TGC 5292
Qy      1909  LysGlnAenHisTyArgProProGlySerProThrCysLeuLeuCysAenCysTyTyPro 1928
Db      5293  AAGGAAATTTACTACAAAGCCCAAGCCAGATCGTTGCTTCCCTGTGACTGTTCGCC 5352
Qy      1929  ThrGlySerLeuSerArgValCysAenProGlnGlyCysProCysTyTyProGly 1948
Db      5353  CGGC--TCCCAAGCCGTGCTGCAAGACAGACACTGGGCACTGTGCGCAAGCTGTGT 5410
Qy      1949  ValIleGlyArgGlnCysAenArgCysAenAenProPheAlaGlyValThrThrAenGly 1968
Db      5411  GTCACTGCGCGCTCAAGGCAACCGCTGTGATTCCTTTCGCGAAGGTCACTCGCTCGGC 5470
Qy      1969  CysGlnValAenTyTyAenSerCysProArgAlaIleGlnAlaGlyTyrTyrThrProArg 1988
Db      5471  TGTGAAGTATCTACAAAGGTGTGTCCAGAGCAATTTGAAGCTGGCACTGTGTGGCCACG 5530
Qy      1989  ThrArgPheGlyLeuProAlaAlaAlaProCysProGlyGlySerPheGlyThrAlaVal 2008
Db      5531  AAGAATTTGGGAGCGACAGCAAGGCTGTATGCGCCAAAGATCCGTGTGTAACCGAGTC 5590
Qy      2009  ArgHisCysAenGlnHisArgGlyTyTyLeuProProAenLeuPheAenCysTyTySerIle 2028
Db      5591  CGGCACTGTGAGTGGGAGAGGCTGTGCTTCCCGCAAGCTCTTCAACTGCACTCTGCGC 5650
Qy      2029  ThrPheSerGlyLeuTyGlyPheAlaGlyValArgLeuGlnArgAenGlySerGlyLeuAenPro 2048

```

```

Db      5651  TCTTTGTGACCTCAAGGCTTGAACGAGAACTGAACCGCAACGACAAATGAGAC 5710
Qy      2049  SerGlyArgSerGlnGlnLeuAlaLeuLeuArgAenAlaThrGlnHisThrAlaGly 2068
Db      5711  GGAACCGGCTCTGCGGCTGGCAAGGCTGTAGAAAGCCGCAAGGAGAAACGACACC 5770
Qy      2069  TyrPheGlySerAenValIleValAlaTyGlnLeuAlaThrArgLeuLeuAlaHisGly 2088
Db      5771  CTCTTGGCAATGATGTGCGACCGCTTACAGGCTTCTGCGCCGATCTTACAGCATGAG 5830
Qy      2089  SerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAenValHisPheThrGlnAenLeu 2108
Db      5831  AGCGCGCAGCGGCTTGTGACCTGGACCGCAACCGAGGCTTAATTTTCATGAGGATGTC 5890
Qy      2109  LeuArgValGlySerAlaLeuLeuAenPheAlaAenTyTySerGlyLeuGlnIleGln 2128
Db      5891  GTCACTACAGGCAAGCGCTCTGCGCCCACTACAGAGCATGTGGGAACAGTCCAG 5950
Qy      2129  GlnThrGlnGlyGlyThrAlaTyTyLeuLeuGlnHisTyTyGlnAlaTyTyAlaSerAlaLeu 2148
Db      5951  CGAAGCGAGGCTGTGCGACGCACTACTGAGGCACTTTCAGGCACTTTCAGCAACGTG 6010
Qy      2149  AlaGlnAenMetArgHisThrTyTyLeuSerProPheThrIleValThrProAenIleVal 2168
Db      6011  GCACGAAATGTGAAGAGCACTATCTGAGGCTTCTGTCACTGTCACTGCAACATGATT 6070
Qy      2169  IleSerValValArgLeuAPLeuAPGlyAenPheAlaGlyAlaTyTyLeuProArgTyGly 2188
Db      6071  CTTCAGATTGACATCTTGCACAAAGCTGAACTTCAAGGCTGCGCCAGGTGCCAAAGTTTGA 6130
Qy      2189  AlaLeuArgGlyGlnGlnProProAPLeuGlnThrThrValIleLeuProGlySerVal 2208
Db      6131  GACATTCAGAAAGAGCTCCCAAGGAGGTGAGTCTCCGCTGCTTCCAGCTGACACC 6190
Qy      2209  Phe-----ArgGlnThrProProValValArg----- 2217
Db      6191  TTCAGCCACAGAGAAAAAGAGCCCTGTGTGTAGGCTGACCAACCGAGAGACTACC 6250
Qy      2218  -----ProAlaGlyProGlyGlnAlaGlnLeuProGlnGlnLeuAlaArgArgGlnArg 2235
Db      6251  CCACGTACCGGACAAACAGAGCCGAGGCTGAGAGGAAACCTCATCAGACAGACAGAG 6310
Qy      2236  ArgHisProGlyLeuSerGlnGlyAlaValAlaSerValIleIleTyTyArgThrLeu 2255
Db      6311  AGACACCCCGAGTGAAGCTGAGCACTTGTGTGCTCCCTGTGTGCATTTACCGGACCTGT 6370
Qy      2256  AlaGlyLeuLeuProHisAenTyTyAenProAPLeuAenTyTySerIleAenValProTyArg 2275
Db      6371  GGTCACTGTGTCTGAACTATGACCCCGACATGCGACGCTCCGACTCTTAAACCGG 6430
Qy      2276  ProIleIleAenThrProValIleSerIleSerValHisAenAPGlnGlnLeuLeuPro 2295
Db      6431  CCTGTACATCAACACCCCGTGTGTGATGTGTATGTGTATCACTGAGGAGAACCCACTGCC 6490
Qy      2296  ArgAlaLeuAenTyTyProValThrValGlnPheArgLeuLeuGlnThrGlnGlnArgThr 2315
Db      6491  AGCTCTGTGAGAGGCTATCTGTGTGAGATTCCTCTGTGTGAGAGCGAAGAAAGCAAGC 6550
Qy      2316  LysProIleCysValPheTyTyAenHisSerIleLeuValSerGlyTyTyThrSer 2335
Db      6551  AAACCTGTGTGTATCTGGAACCACTCTCTGACACTGTGTGAGACTGTGGAGGAGGTGCA 6610
Qy      2336  AlaArgGlyCysGlnValAlaPheArgAenGlyLeuSerHisValSerCysGlnCysAenHis 2355
Db      6611  GCCAAGGCTGTGAATCTTGTGAGAGCCGACCCACCGTCACTTGCAGTGCAGTGCAGCAT 6670
Qy      2356  MetThrSerPheAlaValLeuMetAenValSerArgArgGlnAenGlyGlnIleLeuPro 2375
Db      6671  TCGGCAAGCTGCGGAGTGTCAATGACATTTCCAGAGTGCACAGGAGAGTCTGTGCC 6730
Qy      2376  LeuTyTyThrThrTyTyValAlaLeuGlyValThrIleValAlaLeuLeuLeuThrPhe 2395
Db      6731  CTGAAGATCACTCACTTACGCGCTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6790

```





XX Burner GC, Roush CL, Brown JP;  
XX WPI, 2003-046718/04.  
DR P-PSDB; ABP81741.  
XX

PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
PT (GPCR), useful for diagnosing and designing drugs for treating  
PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
PT cancer or autoimmune diseases

PS Disclosure: Fig 1, 523pp; English.

XX The present invention describes antigenic peptides (I) comprising:  
CC (a) any one of 1601 sequences (see ABP82019 to ABP8619) of 12-24 amino  
CC acids. Also described: (1) an assay for the detection of a particular  
CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity  
CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
CC an antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related diseases, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. ABP42523 to ABP42689 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention.

XX Sequence 11965 BP; 2239 A; 3697 C; 3645 G; 2384 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 11965  
Score: 7836.50 Matches: 1570  
Percent Similarity: 66.71% Conservative: 424  
Best Local Similarity: 52.53% Mismatches: 817  
Query Match: 50.41% Indels: 179  
DB: 25 Gaps: 48

US-09-916-849A-3 (1-2923) x ABP42586 (1-11965)

QY 28 ProLeuLeuGlyAspGln-----ValGlyProCysarg 38  
DB CCTCTTGGGGGGGAGACAGACCTTTGGCAAGAGTATGTCACACAGGGGCTGTCT 759  
QY 39 SerLeuGlySerArgGlyArgGlySer----- 48  
DB 760 CAG---GGGTCCCGGGCTCGGGGAAACAGCTCCCTCCCTTCAAGACTTTTGATTGGC 816  
QY 49 -----GlyAlaCysAlaProMetGlyTyrPleuCysProSerSerLaserxan 64  
DB 817 ACCACGGTCCCAAGCCGGGTCTCTCCACAGGAAAGCTGGAGACGGCTCCCAAAAG 876  
QY 65 LeuTyrPleuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84  
DB 877 ---TGGGACACCGCGCTCTGTGGGAAATTATGGCAACAG-----GAGCAAGGGTC 927  
QY 85 -----ProHisHisAspGlyLeuArgValTyrCysPProGluSerGluLahis 100  
DB 928 AGGGGAGAGAGACCGACAT-----CCGAGCAGAGAAAGAGACG 966  
QY 101 IlePro-----LeuProProAlaProGluLysCys 110

DB 967 CCCCCGGGAGAACTGCTTCCAGGGGCTCGGGAATCGGCCCGAGCTG--GATTCAACA 1025  
QY 111 ProTyrSerCysArgLeuLeuGlyLysGlyHisLeuSerProGluGlyLysLeuThr 130  
DB 1026 CCAAGCAGCGGAGAGACAGCTCTGATCAAGTTCA---GCAACCCCGAGACTCGGACA 1082  
QY 131 LeuProGluGluHisProCysLeuLysAlaProArg-----LeuArgCysGlnSerCys 148  
DB 1083 GCTCCCGAGCGGCGCCCAAGCGCATGCGCTCCGGGGTCTCTTCCGCTGCCGCTTC-- 1139  
QY 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLys 161  
DB 1140 ---CTCCGAGCGCCCGCGGCGCGCTCCCGGAGCTCCCGCCCTCGAAGCCAG 1196  
QY 162 SerProGluGluSerLeuGlyLysArgGlySerAspAlaThrAlaProGluPhe 181  
DB 1197 AAAGTAACTCGGCAACCGGCGACGCTTGTGCGCCGCAACCGCACCCGAGTTT 1256  
QY 182 GlnProProSerTyrGlnAlaThrValProGluGluGlnProAlaGlyThrProValAla 201  
DB 1257 CCGCAGTCAACTACACAGACCTGGTGGCGGAGATGAGCAGCAGCAGCGCGTGTCTA 1316  
QY 202 SerLeuArgAlaIleAspProAlaGluGlyLysAlaGlyValGluLeuGlyThrMetAsp 221  
DB 1317 CGCGTGTGCTCAGAGACCCGAGCGCGGAGCGCGGCGCTTACTCTGCTGCG 1376  
QY 222 AlaLeuPheAspSerArgSerArgGlnPheSerLeuAspProValThrGlyValAla 241  
DB 1377 GCATCAAGAACACCGCTGCTGAGCTGAGCTTTCAGATTCAGACCCGAGCGCTTATC 1436  
QY 242 ThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261  
DB 1437 CGTACGGGAGGAGCTGAGACCGGAGAGATGAGACCTCACTACCTGCTGACCGCG 1496  
QY 262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAsp 281  
DB 1497 CAGGACACCGGGTGGCGCTCTCGGCAACAGATGAGCGCGCTGACAGTACAGTAC 1556  
QY 282 ThrAspAspHisAspProValPheGluGluGlnGluTyrLysGluSerLeuArgGluAsn 301  
DB 1557 CGCAACGACCACTGCGCGGTTTGGAGCAAGCGAGTACCGGAGACCTTCCGAGAAAT 1616  
QY 302 LeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsn 321  
DB 1617 GTGAGGAGGAGCTACCTATCTCAGCTGCGTCCACTGACGCGGAGCGCCGCCCAAC 1676  
QY 322 AlaAsnIleLeuTyrArgLeuLeu-----GluLysSerGlySerProSerGluVal 339  
DB 1677 GCACACCTGCGCTACCGCTTCTGTGGGCGCCGACGCTGCGCGCTGCGCGCC 1736  
QY 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluLys 359  
DB 1737 TTGAGATGATTAACAGCTCCGCTCAGCTCAGACACGCGCCAGAGTGCACCGAGAC 1796  
QY 360 ValGluSerTyrGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379  
DB 1797 ATGAAAGCTATGAGCTGTGTGGTGAAGCCAGACAGGCGCAGAAACCGGGCGCGC 1856  
QY 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAspAspAlaProGluPhe 399  
DB 1857 TGGCCTGCTGGCGCGTACATATCTGTCTAACAAGAACGAAATGCTCTCACTTC 1916  
QY 400 SerGluArgTyrValValGlnValArgGluAspValThrProGlyAlaProValLeu 419  
DB 1917 AGCGAAGAGCGCTACGTTGGGCGAGGTGCGGAGATGTGGCCCCCACAAGTGTGCTG 1976  
QY 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValValHisTyrSerIleMet 439  
DB 1977 CGGTCAACGCGCATGACCGGAGCAAGAGCGCAACGATTTGGTCACTACATCATC 2036  
QY 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValAlaAspValVal 459  
DB 2037 AGTGGATATAGCCGTGACACTTTGCAATGACGCTTCACTGGCGAGATCAGGTGCTG 2096

Qy 460 SerProleuAspTyrGluThrThrLeuGluTyrThrLeuArgValArgAlaGlnIleAspGly 479  
Db 2097 GCACCTTGAGACTTCGAGGAGAGAGAGATGCCCTTGCGCATCAGAGGCCAGATGCT 2156  
Qy 480 GlyArgProProleuSerAsnValSerGlyLeuValThrValGlnValIleAspIleAsn 499  
Db 2157 GGGCGGCGACCGGCTGTCACCAACACCGGCGTGGCCCATCCGAGGTGGAGCATCAAT 2216  
Qy 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
Db 2217 GACACATTCCTATTGTCGAGCAGCCCTCCAGATTCCTGCTTGGAAAAATGCTCCC 2276  
Qy 520 LeuGlyTyrLeuValIleuHisValGlnAlaIleAspAlaAspAlaArg 539  
Db 2277 TTGGGTCACTCAGTCATCCATTCAGCGAGTCGATCAGACATCGGAGGAATGCCAG 2336  
Qy 540 LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
Db 2337 TTGGAGTACCTCCCTAACCTGGGTGGCACCTGATCTCTTTGTGATAAACAGCGCAC 2396  
Qy 560 GlyTyrIleSerValAlaIleGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
Db 2397 GGCTGGGTCTCTGTGAGGTGCTCCCTGACCGTGAAGTCTGTGGAGCATTCCTTTGGT 2456  
Qy 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThr 599  
Db 2457 GTGGAGGCTCGAGACATGCGTCAACCCCACTGCTGCTCAGCGAGTGCACCGTGA 2516  
Qy 600 ValLeuAspValAsnAspAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
Db 2517 GTGTGACCGTATGACAAATCGGCTGAGTTCACAAATGAAGGAGTTCACATCAAGCTG 2576  
Qy 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspAspAla 639  
Db 2577 AATGAGATCAAGCTGTGGGACCAAGTGTGAGCTGTGGAGCATTCCTTTGGT 2636  
Qy 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
Db 2637 AACGTGCATCAGCTACAGATCAGACGCGGACACCCGGAATGCTTTGCGATCAGC 2696  
Qy 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProleuAspTyrTyrGluGluArg 679  
Db 2697 ACCGAGGGGGTGTGGTCTGTGATCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTG 2756  
Qy 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
Db 2757 TACTTCAAGCTGACTACATGCTGATCGACCGTCCCTTCAATGATCTCTATGTCAC 2816  
Qy 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrVal 719  
Db 2817 ATCAACATCAAGATGCAACATCACTGACCGGCTTTTCAAGATGCCACTACTCAGTG 2876  
Qy 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
Db 2877 AGTGTGATAGATGATCGGCAATGGGATGAGCACTAATGTGATCAATGCTCTGATGAT 2936  
Qy 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
Db 2937 GACGTGGGTAGATGCTCGATCACTATCTCTGAGAGACAACTGCCCGCAGTTCCGC 2996  
Qy 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
Db 2997 ATTGATCAAGATCAAGAGCATTAATCAAGGCCCATTAAGCTATGAGACAGAGTG 3056  
Qy 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnPheSerAspThrThr 799  
Db 3057 ACCATCACTGAGTGTGATCAAGCTCGGAGCAATGGCATCCACAGAGGAGACACTACT 3116  
Qy 800 TyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
Db 3117 TATGTGAGGTGATGTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 3176

Qy 820 TyrGlnGlySerValTyrGluAspValProPheThrSerValLeuGlnIleSerAla 839  
Db 3177 TATACAGGGCTGTGCTGAGAGATGCCCACTTTACAGAGTCTCTGAGATCTCAGCC 3236  
Qy 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyAspAsp 859  
Db 3237 ACTGACCGGAGTGCATATCCCAATGCGCGGTCCAGTACATCTTCCAGAAATGTGAAGAT 3296  
Qy 860 GlyAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeu 879  
Db 3297 GGGAGATGAGATTTTACATTAAGCCCATCTTGAAATGTCTGATCAAGTAAAGCGGCTA 3356  
Qy 880 AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspArgGlyMetPro 899  
Db 3357 GACCGGAGGAGCATATCAAGTATGAGTTGATGCTGCTACGAGTGGACAGAGGTGGCC 3416  
Qy 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnPro 919  
Db 3417 CCACTCCGAGCTCCAGTCAATCAAGGTGAGTGGAGATGTGAACGACATGACACT 3476  
Qy 920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 3477 GTCTTCCAGCTGAGGAGTTGAGGTGCGGTGAAGAGATGATGATGTGGGCTCAAGTG 3536  
Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3537 GTGGCCCAATCACTGAGTGAAGTGAACCCCTTAAGAGGCCCAATTAATATGATCAAG 3596  
Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
Db 3597 ATCGTGAGGGAGAAATCCCTGAGCTGTTCCAAATGACATCTCTCTCGAATCACTGACG 3656  
Qy 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValValIleGlnAlaThr 999  
Db 3657 GACATCATTAACCTTACATATAGAGCTGCCCAAGATATGATATGTGAGAGGCCACA 3716  
Qy 1000 SerAlaProleuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3717 TGTGCTCTTTGGTGAAGCGGGCACTGTGACAGTCCCGCTGGTGAACAGATATGACAA 3776  
Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3777 AGCCCTGTCTCAACACTTCAGATTCCTTCAACATATGATATCAACCGTTCAGAC 3836  
Qy 1040 SerPheProGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3837 ACCCTCCGCTGGGCAATATTTGGGCGCATCCAGCTATGATGATCCGATGCTCCAGCAC 3896  
Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
Db 3897 CTCTTCTACTCTTTGAGCGGTGGCAATGAGCTGAGCTGCTGTGATGATCAACAGCACT 3956  
Qy 1080 GlyGluLeuValLeuSerArgAlaLeuAspAsnAsnArgProleuGlnAlaIleMetSer 1099  
Db 3957 GGGAGCTGCACTACGCGGAAGCTACAAATTAACCGCCACTGCTGCTCCATGTG 4016  
Qy 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 4017 GTGACGTCAAGATGCGCTGACAGCGTGAACGGGCGAGTGTGTGCGGTGCTGATC 4076  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
Db 4077 ATCAACGAGAGATGTGTGCCCAACAGCTGACCGTGGCTTGAAGAACATGTGGCAGAG 4136  
Qy 1140 ArgPheLeuSerProleuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 4137 CGCTTCTCTCAACCGCTGTGGCGGCTTCTCTCAAGGGGTGCTGCGGTCTCGCTAG 4196  
Qy 1160 ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
Db 4197 CCGCGTGAAGAGTCTCTTCAATCTCAACATCCAGAACAGACAGAGAGCTA---GGGGCAGCC 4253  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyPro----- 1197

```
Db 4254 GGGCTAAAGTGAAGTTTCTCGGCGTAGCTCCAGTGGGGCGGGCGCGCTGCAGGG 4313
Qy 1198 ProPheLeuProSerGluAapLeuGlnGluArgLeuTyLeuAapSerLeuThr 1217
Db 4314 CCTGGTTACAGCTCCGAGAGAGCTGCAGAGACATTGACGTGGCGGGCGGCGCTGGG 4373
Qy 1218 AlaLeuSerAlaGlnArgValLeuProPheAapAapAapAapAapAapAapAap 1237
Db 4374 GCTCGCTCCCTCGTCCAGCTGACCTGCTTCAAGACAGTGGCTGCAGAGAGCTGCT 4433
Qy 1238 GlnAapTyMetArgCysValSerValLeuAapPheAapSerSerAlaProPheAla 1257
Db 4434 GAGAACTAACATGAATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4493
Qy 1258 SerSerSerValLeuPheAapProLeuAapProValGlyLeuArgCysArgCysPro 1277
Db 4494 TGGGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4553
Qy 1278 ProGlyPheThrGlyAapTyCysGlnThrGlyValAapLeuCysTySerArgProCys 1297
Db 4554 CCGGATTCACGGAGACATTTTGGCAGACCGAGCTGACCTGCTGCTGCTGCTGCTGCT 4613
Qy 1298 GlnProHisGlyArgCysArgSerArgGlnGlyTyThrCysLeuCysArgAapGly 1317
Db 4614 CCGAACGGCGAGAGCTGCGCGCGCGCGAGAGAGCTACACGTCGCTGCGCGCGCGC 4673
Qy 1318 TyThrGlnGlyHisGlyCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCys 1337
Db 4674 TTCACCGGAGAGAGAGCTGCGAGCTGCACACCGAGCGCGCTGCTGCTGCTGCTGCTG 4733
Qy 1338 LysAapGlnGlyTyThrCysValAapLeuValGlyTyPheLysCysAapCysProSer 1357
Db 4734 CCGAACGGCGAGAGCTGCGACCGACCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTG 4793
Qy 1358 GlnAap---PheGlnLysProTyCysGlnValThrThrArgSerPheProAlaHisSer 1376
Db 4794 GCGGCGCGCTGAGAGCGCGCGCTGCGAGGTGCTGCGCGCTGCTGCTGCGCGAGTTCG 4853
Qy 1377 PheLeuThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAla 1396
Db 4854 TTCGTCATGTTTCGCGGCTGCGGCGAGGATTCACCTTACGCTGCTGCTGCTGCTGCTG 4913
Qy 1397 ThrLysGlnArgAapGlyLeuLeuLeuTyArgGlnArgPheAapGlnLysHisAapPhe 1416
Db 4914 ACAGTGCAGACAGCGGCTGCTGCTTTCACACGGCGCTGACACAGAGACACACCTTC 4973
Qy 1417 ValAlaLeuGlnValIleGlnGlnValGlnLeuThrPheSerAlaGlyLysSerThr 1436
Db 4974 CTGGCCCTGGAACTGCTGCTGCTGCGCAAGTGCAGCTTCAATATTCACCGGTGAATCCAC 5033
Qy 1437 ThrThrValSerProPheValProGlyValSerAapGlyGlnThrPheThrValGln 1456
Db 5034 ACCGAGTCAAGCCCAAGTTCAGGGGCTTGAAGTGAAGGCGCATGACATGACAT 5093
Qy 1457 LeuLysTyTyTyAapLysProLeuLeuGlnGlyGlnThrGlyLeuProGlnGlyProSerGln 1476
Db 5094 CTGAATATCTACAAACAGCCCGCAGATGCTTGAAGGGGTGCACAGGCGCCCTCCACAG 5153
Qy 1477 GlnLysValAlaValValThrValAapGlyCysAapThrGlyValAlaLeuArgPheGly 1496
Db 5154 GACAAAGTGGCTGTGCTGAAGCGTGAATGTGATGTGGCGGTGCTTGCAGTTGGT 5213
Qy 1497 SerValLeuGlnAapTySerCysAlaAlaGlnGlyTyThrGlnGlyGlySerLysLysSer 1516
Db 5214 GCTGAGATTGGCACTACTCATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5273
Qy 1517 LeuAapLeuThrGlyProLeuLeuGlnGlyValProAapLeuProGlnSerPhePro 1536
Db 5274 CTGAGACCTGACCGGCGCTTCTTCTGAGAGGTGCTCCCAACCTCCCGACATCTCC 5333
Qy 1537 ValArgMetArgGlnPheValGlyCysMetArgAapLeuGlnValAapSerArgHisIle 1556
Db 5334 GTATCCCATTAAGACTTCATGCGCTGTATGCGGAGCTGCACATTGATGCGCGGAGTGC 5393
Qy 1557 AapMetAlaAapPheIleAlaAapAapGlyThrValProGlyCysProAlaLysLysAap 1576
Db 5394 GACATGGCGGCTTTTGTGCGAAATATATGGACCATAGGAGCTTGCACAGCCATAC 5453
Qy 1577 ValCysAapSerAapThrCysHisAapGlnGlyTyThrCysValAapGlnThrAapAap 1596
Db 5454 TTTTGTACTCAGGCGCCCTGCAGAAACAGTGGCTTCTGCTGCGAGCGCTGGGGCAGCTTC 5513
Qy 1597 SerCysGlnCysProLeuGlnPheGlyLysSerCysAlaGlnGlnLeuAlaAapPro 1616
Db 5514 AGCTCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5573
Qy 1617 GlnHisPheLeuGlnSerSerLeuValAlaThrHis---GlyLeuSerLeuProIleSer 1635
Db 5574 CACCATTTCCGCGGACAGGACACTGAGCTGGAATTGGAAAGTGAAGTGAAGTGTGCT 5633
Qy 1636 GlnProTyTyTyLeuSerLeuMetPheArgThrArgGlnAlaAapGlyValLeuLeuGln 1655
Db 5634 GTGCGATGGTACCTGGGCTGGCATTTGCGACACGGGCAACGAGGGGGTCTGATGACA 5693
Qy 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuAapGlnGlyHisValMetLeu 1675
Db 5694 GTGACGCTGGGCGCACACAGACGCTCTTGGCAGTGAATCGGGGGTTACTGTCTGTG 5753
Qy 1676 SerVal---GlnGlyTyThrGlyLeuGlnAlaSerSerLeuAapGlnProGlyArgAla 1694
Db 5754 ACAGTGCACAGAGGCTGCGGCG---CGTCTTCCCATCTCTTCTGACAGTGAATGCTG 5810
Qy 1695 AapAapGlyAapThrPheHisAlaGlnLeuAlaLeu---GlyAlaSerGlyLysPro 1712
Db 5811 AGTATGCGCGGCTGACAGATCTGCGCTGAGTTCAGAGAGAACACAGTGGCGCGCG 5870
Qy 1713 GlnHisAlaIleLeu---SerPheAapTyGlnGlnGlnArgAlaGlnLysLeu 1730
Db 5871 GCGCACATGCTTATGCTCTCATGCTGCTGACCTTACCTTCCAGACACCATGCGGCTG 5930
Qy 1721 GlnProArgLeuHisGlyLeuHisLeuSerAapIleThrValGlyLys---ProGly 1749
Db 5931 GGGAGTGAAGCTGCAGGCGCTGAAGGTAAAGACAGCTCCAGTGGAGGCTGCGCCCGG 5990
Qy 1750 ProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAap 1769
Db 5991 AGTGCAGAGAGGCTCTCCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6050
Qy 1770 ThrProGlnGlyValAapSer---LeuAapProSerHisGlyLysSerIleAapValGln 1788
Db 6051 ACACCTCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6104
Qy 1789 GlnGlyCysSerLeuProAapProCysAapSerAapProCysProAlaAapSerTyCys 1808
Db 6105 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6164
Qy 1809 SerAapAapThrAapSerTySerCysSerCysAapProGlyTyTyTyGlyAapAapCys 1828
Db 6165 CCGGACCTCTGCGACAGCTTCTTCTGACCTGCGACGCGACAGGTATCTAGGCGCCAGGCTG 6224
Qy 1829 ThrAapValCysAapLeuAapProCysGlnHisGlnSerValCysThrArgLysProSer 1848
Db 6225 GTGAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6284
Qy 1849 AlaProHisGlyTyTyThrCysGlnCysAapProAapThrLeuGlyProTyCysGlnThr 1868
Db 6285 GCGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6344
Qy 1869 ArgIleAapGlnProCysProArgGlyTyTyTyPheGlyHisProThrCysGlyProCysAap 1888
Db 6345 AGGATGACACAGAGTGCACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6404
Qy 1889 CysAapValSerLysGlyPheAapProAapProAapLysThrSerGlyLysCysHisCys 1908
Db 6405 TGTGATGTTCAAGAGTTTGAATCCCACTGCACAGACAAATGGGCACTGCTGCTGCTG 6464
```

QY 1909 LysGluAsnHisArgProGlySerProThrCysLeuLeuCysAspCysTyrPro 1928  
 DB 6465 AAGAGATTCACATACGACCGCGGGGAGTGAAGCTCTGGCTCCCATGTGACTCTACCT 6524  
 QY 1929 ThrGlySerLeuSerArgValCysAspProGluAspGlyGlnCysProCysLysProGly 1948  
 DB 6525 GTGGGGCTCCACCTCGCGCTCATGTGACCCACACGCGGGGAGTGGCTCTCCCGCCAGGA 6584  
 QY 1949 ValIleGlyArgGlnCysAspArgCysAspAsnProPheAlaGluValThrThrAsnGly 1968  
 DB 6585 GCCCTTGACCGGACGACGACGCTGACAGTCCCTTCGACAGAGGACAGCCAGCGGC 6644  
 QY 1969 CysGluValAsnTyrAspSerCysProArgAlaIleGluAlaGlyIleTyrTyrProArg 1988  
 DB 6645 TGCAGGGTGTCTATGATGCTGCTCCCTAAGTCCCTGATGCTGTGTGTGGCCAG 6704  
 QY 1989 ThrArgPheGlyLeuProAlaAlaAlaProCysProLysGlySerPheGlyThrAlaVal 2008  
 DB 6705 ACAAGTTTGCGCTGCGCCACAGTGCCTGTCCCGGGGGCGCTGGGTGTGTGTG 6764  
 QY 2009 ArgHisCysAspArgLysArgGlyTyrPheLeuProProAsnLeuPheAsnCysThrSerIle 2028  
 DB 6765 CGGCTGTGTATAGGCGCCAGGGTTGGCTGAGACCCGACCTTTCACATCTTACCTCCCT 6824  
 QY 2029 ThrPheSerGlyLeuLysGlyPheAlaGlyLeuGlnArgAsnGlySerGlyLeuAsp 2048  
 DB 6825 GCCTTTCAGAGCTCACTGTGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 6884  
 QY 2049 SerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisIleThrAlaGly 2068  
 DB 6885 ACATAGGAGGCAAGAGGCTGCTGACGCGGCTACGGGAGGTGACGTGCGCCACTGACAC 6944  
 QY 2069 TyrPheGlySerAspValLysValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlu 2088  
 DB 6945 TATTTTAGCCAAAGATGTTCAAGTCACTGCCGCTGTGTGGCCCACTGTGGCTTGCAG 7004  
 QY 2089 SerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnLeu 2108  
 DB 7005 AGGCATCAGAGGGCTTGGGGCTGACAGCACACAGGATGCCCACTTCATGAGAAATCTG 7064  
 QY 2109 LeuArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisIleThrGluLeuIle 2127  
 DB 7065 CTGGGGCGGCTGTGACTGTGCTGCCACAGACACAGGGACCTTGGGGCGGCTGGGG 7124  
 QY 2128 GlnGlnThrGlnGly-----GlyThrAlaThrPheLeuGlnHisIleTyrGluAlaTyr 2144  
 DB 7125 CAGGGGGCCCTGGGGGCTCCCGACGACGGCGGGAATGGTGAAGGACCTGGAAGATAT 7184  
 QY 2145 AlaSerAlaLeuAlaGlnAsnMetArgHisIleThrTyrLeuSerProPheThrIleValThr 2164  
 DB 7185 GCGAGCCACACTCGGACAGAAATATGAATCAATACCTGAATCCCATGGGGCTGTGTACG 7244  
 QY 2165 ProAsnIleValIleSerValAlaArgLeuAspLysGlyAsnPro-----AlaGlyAla 2182  
 DB 7245 CCTTAATACAGCTCAGCATATGACCGGATGAGACCCCAAGTCTCCCGGGGGGGCCGT 7304  
 QY 2183 LysLeuProArgTyrGluAla-----LeuArgGlyGlnGlnProProAspLeuGluThr 2200  
 DB 7305 CGCATACCTCGCTACCATAGCACTCTTTCGAGGCAAGATGCTGGATTCCTCAACAC 7364  
 QY 2201 ThrValIleLeuProGlyLeuSerValPheArgGlyThrProProValValAlaArgProAlaGly 2220  
 DB 7365 CATGT 7424  
 QY 2221 ProGlyGlnGlnGlnProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2240  
 DB 7425 AGCAGCATAGAAATCCACACCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7484  
 QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 DB 7485 GAGCTGT 7544

QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
 DB 7545 GCCAATTCAGACAGAAAGCCGAGGTGACAGGCTTCTCCAGAAACCCCTGCAATGACTCC 7604  
 QY 2281 ProValIleSerIleSerValHisAspArgGlyGlnLeuLeuProArgAlaLeuLeuLys 2300  
 DB 7605 CCGGT 7664  
 QY 2301 ProValIleValIlePheArgLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2320  
 DB 7665 CCATCAGCTGAGATTTGCTGCTGCTGACAGACAGGAAATGAGACAGGCAATCTGTGTG 7724  
 QY 2321 PheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGlyCysGlu 2340  
 DB 7725 CAGTGGAGCCACCTGGCTGGGAGAGCATGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7784  
 QY 2341 ValValPheArgAsnGlnSerIleValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
 DB 7785 CTGGTGCACAGAAAGGGTCCCAAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7844  
 QY 2361 ValLeuMetAspValSerArgArgGlu-----AsnGlyGlnIleLeuProLeuLysThr 2378  
 DB 7845 GTCTCATGATGATGCTCTCCCGCTGAGAGCTGACAGGCGGACCTGAGCTGTGTGTG 7904  
 QY 2379 LeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeu 2398  
 DB 7905 TTCAACCAAGT 7964  
 QY 2399 ThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaIle 2418  
 DB 7965 CTGAGCTTCGAGGCTCAAGTCAATGTGCTGGAGTCAATGACCAATGTGCAGCGGC 8024  
 QY 2419 LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAla 2438  
 DB 8025 CTGGGGGTGAGAGGTCTCTCTCGTGGGGATTCACAGAACCCACAAATCAGCTGTG 8084  
 QY 2439 CysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerThrAlaLeu 2458  
 DB 8085 TGCATGCAAGTGCACATCTCTGCACTTCTCTCTGACGACCTTCCCGTGGCTTTC 8144  
 QY 2459 LeuGluAlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyPro 2478  
 DB 8145 GTGCAGGGCTGACCTTCAACCGCATGACAGTGAACACGCACTGACCGCGGCGCC 8204  
 QY 2479 MetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaVal 2498  
 DB 8205 ATGGCTTCTACCATGCTCTGGGGCTGGGGCTGCCGTGTGTGTGTGTGTGTGTGTGTGT 8264  
 QY 2499 GlyLeuAspProGlyGlyTyrGlyAsnProAspPheCysThrPheSerIleTyrAspThr 2518  
 DB 8265 GGGCTGAGCCCTGAGGGCTATGGAACTCTGACTTCTGTGATTCAGTCCACAGGCC 8324  
 QY 2519 LeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyr 2538  
 DB 8325 CTGATGGAAGCTTGT 8384  
 QY 2539 IleLeuAlaAlaArgAlaSerCysAla-----AlaGlnArgGlnGlyPheGlyLysGly 2557  
 DB 8385 CTCCTGCTGCGCGCACATCTGTCTCCACAGGGCAGAGGAGGAGCC-----AAGAAAGACC 8438  
 QY 2558 ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrThrPheLeu 2577  
 DB 8439 TGTGACATACCTTGTGAGCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8498  
 QY 2578 LeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHisTyrThrPheAlaThrCys 2597  
 DB 8499 TTGGGCTCTGCGAGTCAACACAGCATCTTCACTTCACTTCACTTCACTTCACTTCACTT 8558  
 QY 2598 AsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGluValArg 2617  
 DB 8559 TGGGCTTCAGGGCTGCGGT 8618  
 QY 2618 LysAlaLeuLysLeuAlaCys---SerArgLysProSerProAspProAlaLeuThrThr 2636



(e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle, ovary, testis and uterus), neurological disorders (e.g. epilepsy, stroke, ischaemic cerebrovascular disease, Alzheimer's disease or Pick's disease), disorders of vesicular transport (e.g. cystic fibrosis, diabetes mellitus, Grave's disease, or gastric and duodenal disorders), disorders (e.g. ulcerative colitis, or gastric and duodenal disorders), autoimmune diseases (e.g. allergic reactions, autoimmune hemolytic anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic and protozoal infections. The NOVX proteins can be used as immunogens to produce antibodies and as vaccines. The NOVX nucleotide sequences may be used in chromosome mapping, identifying individuals from minute biological samples (tissue typing), and in forensic identification of a CC located on chromosome 3q21.3-4.

XX Sequence 12348 BP; 2287 A; 3911 C; 3696 G; 2446 T; 8 other;

# Alignment Scores:

Pred. No.:	Length:	12348
Score:	7815.50	1566
Percent Similarity:	66.14%	Conservative: 428
Best Local Similarity:	51.94%	Mismatches: 819
Query Match:	50.28%	Indels: 203
DB:	24	Gaps: 47

US-09-916-849a-3 (1-2923) x AB082327 (1-12348)

```

QY 28 ProleuLeuGlyAapGln-----ValGlyProGlyAapG 38
DB 416 CTTCTTCGGGCGGAGAGACCTTTCAGAAAGAGTACTGTCTACAGGAGCTCTGTCT 475
QY 39 SerLeuGlySerArgGlyArgGlySerSer-----48
DB 476 CAG---GGGTCCCGGGCTCGGGAGAACAGTCCGCCCTCCCTTCAGACTTTTATTCGCG 532
QY 49 -----GlyAlaCyValaProMetGlyTyrLeuCySerProSerSerAlaSerAsn 64
DB 533 ACCACGCTCCCAACCGCGTCTCTCCAGCGAGACGCTCCGCGAGACGCTCCCGCAAAAGAG 592
QY 65 LeuTyrLeuTyrThrSerArgCyValaAapAlaGlyThrGluLeuThrGlyHisLeuVal 84
DB 593 ---TGGGACCGCGCGCTCTCTGGGAGATTATGGGCAACAG-----GAGCAAGGGGTC 643
QY 85 -----ProHisAapGlyLeuAryValTyrCySerProGluSerGluAlaHis 100
DB 644 AGGCGAGAGAGCCAGCAT-----CCGAGCAGAAAGAGCAG 682
QY 101 IlePro-----LeuProAlaAapGluGlyCys 110
DB 683 CCCCCCGGAGACTGTCTTCAGAGGCTTCGGAGTCTGGCCCGGAGCTG-GATTTCAGCA 741
QY 111 ProTyrSerCyValaGlyLeuGlyTyrGlyHisLeuSerProGluGlyLeuThr 130
DB 742 CCAAGCAGCGGAGAGACAGCTCTCTGATCAGATTCAG---GCACCCCCGGAGCTTCGAGCA 798
QY 131 LeuProGluGluHisProCyValaAapAlaProArg-----LeuArgCySerGlnSerCys 148
DB 799 GCTCCCGAGCGGCGCCCAAGCGATGGCTCCGGGGGTCTCTTCGCTGGCGGCTTC---855
QY 149 LysLeuAlaGlnAlaProGly-----LeuAryAlaGlyGluArg 161
DB 856 ---CTCCCGAGCGCGCCCGGCGCGCTCCCGGAGACTCCCGGCGCTCTGAGAGCAGG 912
QY 162 SerProGluGluSerLeuGlyTyrArgArgGlyValaAapAlaAapAlaProGluPhe 181
DB 913 AAATGAATCTCGGAGAACCGGAGACGCTTCTGTGCGCCGCGCAAAACCGGACCGGACTTT 972
QY 182 GlnProProSerTyrGlnAlaThrValProGluAapGlnProAlaGlyThrProValAla 201
DB 973 CCGAGATCACTACAGACCGCTGTGCGGAGATGAGCAGCAGGACCGCGGCTCTCTA 1032
QY 202 SerLeuAryAlaIleAapProAapGluGlyGluAlaGlyArgGluGluTyrThrMetAap 221
  ::: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 1033 CGGTGGTGTCTCAGGACCCGAGACCGCGGAGGCGGCTGACTACTGCTGGCG 1092
QY 222 AlaLeuPheAapSerArgSerAapGlnPheSerLeuAapProValThrGlyAlaVal 241
DB 1093 GACTATATACAGCGCGCTCGGTGAGCTGTTCATGATGACCCGAGAGCGGCTTATTC 1152
QY 242 ThrThrAlaGluGluLeuAapArgGluThrLysSerThrHisValaPheAryValThrAla 261
DB 1153 CGTAGCGCGGAGCTCTTGAGCCGAGAGACATGAGAGCTCATCACTGCTGGTGCAGCGG 1212
QY 262 GlnAapHisGlyMetProAapArgSerAlaLeuAlaThrLeuThrLysLeuValThrAap 281
DB 1213 CAGAGACAGGGTGTGGCGGCTCTCGGCAACAGATGAGTGGCGGTGACATGACGAGC 1272
QY 282 ThrAapAapHisAapProValPheGluGlnGluTyrLysGluSerLeuAryGluAan 301
DB 1273 CGGACAGACCTCGCGCGGTCTTTCAGAGAGCGGAGTACCGGAGACCTTTCGAGAT 1332
QY 302 LeuGluValaGlyTyrGluValLeuThrValaAryAlaThrAapGlyAapAlaProAan 321
DB 1333 GTGAGAGAGGGGTACCTTATCTCTGAGCTGCGTGCATGACGCGGACCGCCCAAC 1392
QY 322 AlaSerHisLeuTyrArgLeuLeu-----GluGlySerGlyLysProSerGluVal 339
DB 1393 GCCAACCTCGCTTACCGCTTCTGTGAGCCGAGCTCGCGGCTGACCTCCCGCGCC 1452
QY 340 PheGluIleAapProAapArgSerGlyValIleAryThrArgGlyProValAapArgGlu 359
DB 1453 TTGAGATTTGATTCAGAGCTCCCGCTCATCAGACCAAGGCGCGAGTGAACCGGAGCAC 1512
QY 360 ValGluSerTyrGluLeuThrValGluAlaSerAapGlnGlyAryAapProGlyProArg 379
DB 1513 ATGAAAGCTATGAGGTGTGTGTGAGAGCCAGCAGCAGAGGCAAGAGACCCGCGCGCC 1572
QY 380 SerThrThrAlaAlaValPheLeuSerValGluAapAapAapAapAlaProGlnPhe 399
DB 1573 TCGGCACTGTGCGGCTACATCTGTGTCTGTAGAGAGAAACAGACATCTCTCCAGTTC 1632
QY 400 SerGluLysArgTyrValValGlnValArgGluAapValThrProGlyValaProValLeu 419
DB 1633 AGGAGAAACCGCTACCGTGGCGAGAGTGGCGGAGAGTGTGGCGCCCAACAGTCTGTGCTG 1692
QY 420 ArgValThrAlaSerAapArgAapLysGlySerAapAlaValaHisTyrSerIleMetC 439
DB 1693 CGGTGACCGGCACTGACCGGAGCAAGCGCAACGATGTGTGATCAACAACATCACC 1752
QY 440 SerGlyAapAlaAryGlyGlnPheTyrLeuAapAlaGlnThrGlyAlaLeuAapValVal 459
DB 1753 AGTGGCAATGACCGTGTGACACTTTCGATCGAAGAGCTCACTGCGAGATCCAGGTGTG 1812
QY 460 SerProLeuAapTyrGluThrThrLysGluTyrThrLeuAryValaAryAlaGlnAapGly 479
DB 1813 GCACCTGTGACTTCGAGCAGAGAGAGAGATGTGCTTCGATCAAGCGCGAGATGCT 1872
QY 480 GlyArgProProLeuSerAapValSerGlyLeuValThrValGlnValLeuAapIleAan 499
DB 1873 GGCCTGGCAGCCCGCTGTCCAAACAACGCGGCTGCGGAGATCAGGTGTGTGACATCAAT 1932
QY 500 AapAapAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519
DB 1933 GACCAATCTCTATTTTGTTCAGACGCGCTTCGATCAAGTTCGTCTGTGAAATAGCTCTCC 1992
QY 520 LeuGlyTyrLeuValLeuHisValGlnAlaIleAapAlaAapAlaGlyAapAapAary 539
DB 1993 TTGGGTCACTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2052
QY 540 LeuGluTyrArgLeuAlaGlyValaGlyHisAapPheProPheThrLysAanGlyThr 559
DB 2053 TTGGAGTACTCTCTTACTGTGTGTGCACTGTATTAATCTTTGTGTATTAACGCGCACT 2112
QY 560 GlyTyrIleSerValaIleAapAlaGluAapArgGluGluValaAapPheTyrSerPheGly 579
DB 2113 GGTGTGTGTCTCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2172

```



QY 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 DB 2173 GTGGAGGCTCGAAGCATGCTCAACCCCACTCTCTCTCAACGATGACACCGGACT 2232  
 QY 600 ValLeuAspValAlaAspAsnProThrPheThrGlnProGluThrValArgLeu 619  
 DB 2233 GTCTGAGCGTTATGCAATCGCCCTGAGTTCACATGAGGATCAACCTACGACTG 2292  
 QY 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspAspAla 639  
 DB 2293 AATGAGAGTGCACCTGTGGGACAGTGTGTGAGCGGAGACCGAGTACCGTGTATGCC 2352  
 QY 640 HisSerValIleThrTyrglnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 DB 2353 AACAGTGCATCAGCTACCAAGATCAAGCGGCGCAACCGGAGATGGCTTGGCATCAGC 2412  
 QY 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleuLeuGluArg 679  
 DB 2413 ACCAGGGGGGTGTGGTGTGAGTGTGACTGTGCTGCTGCACTGAGCTACAGAGCAAGCC 2472  
 QY 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
 DB 2473 TACTTCAAGCTGTACTTAAGTCACTGACCGTCCCTTCATGATCACTGCTATGTGAC 2532  
 QY 700 ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrVal 719  
 DB 2533 ATCAATCATCAGATGCCAACAATCACTCAATCGGCGGTCTTCAAGTCCCACTACCTCAG 2592  
 QY 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 DB 2593 AGGTGAATGAAGATGGCCAAATGGGTAGCAACATGTGTGTCAATGTCCTCTGATGTAT 2652  
 QY 740 AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 DB 2653 GACGTGGGTGAGAAATCTGTATCACTTATCTCTGAGAGCAACCTGCCCAATTTCCGC 2712  
 QY 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 DB 2713 ATTGATGACAGACTCAGAGGACCAATTATCAAGGCCCATTTAGATGAGGACGAGTG 2772  
 QY 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerThrAspThr 799  
 DB 2773 ACCATACACCTGCTATACAGCTCGGAGCAATGGATCCCAAGAGGACAGACATCTACT 2832  
 QY 800 TyrLeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 DB 2833 TATGTGAGGTATGTCTCATGACGTGAATGACATGCTCCACAAATTTGTGGCTCCAC 2892  
 QY 820 TyrGlnGlySerValTyrGluAspValProProPheThrSerValIleuGlnIleSerAla 839  
 DB 2893 TATACAGGCTGTGCTCTCAGAGATGCCCACTTTACACAGTGTCTGACAGATCTCAGCC 2952  
 QY 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnIleGlyAspAsp 859  
 DB 2953 ACTGACCGGAGTGTCTATGCCAATGGCCGCTGCTCAGTACATTTCCAGAAATGGTAAAGT 3012  
 QY 860 GlyAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgLeu 879  
 DB 3013 GGGGATGAGATTTTATCAATGAGCCCACTCTGGAATGTCTCGTACAGTAAAGCGGCTA 3072  
 QY 880 AspArgGluAsnValAlaGlnTyrValIleuAspAlaTyrAlaValAspIleGlyMetPro 899  
 DB 3073 GACCGGAGGAGATTCAGTATGAGTGTGACTGCTCAGTGTGACAGAGTGTGCGCC 3132  
 QY 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
 DB 3133 CCACTCCGAGCTCAGTCACTTACAGAGTATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3192  
 QY 920 ValPheGlnGluAspGluPheAspValPheValGlnGluAsnSerProIleGlyLeuAla 939  
 DB 3193 GTCTTCCAGCTGAGAGATTTTGAAGTGTGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3252

QY 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 DB 3253 GTGGCCAGATATCTGCACTGAGACCTTACAGAGGCCCAATGGCCATTAATATGACAG 3312  
 QY 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 DB 3313 ATCTGAGGGGAGAAATCCCTGAGCTGTTCAAATGAGACATCTTCTCTGAGAGTACGAG 3372  
 QY 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThr 999  
 DB 3373 GCACTCATTTGACCTTATGAGTGTGCTGCGCAAGAAATGATGTGTGTGTGAGCGCA 3432  
 QY 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 DB 3433 TCTGCTCTTGTGTGAGCGGGCACTGTGACGCTCCGCTGTGTGACCAAGATGACAC 3492  
 QY 1020 ProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
 DB 3493 ACCCTGTGCTCAACAATTCAGATCCCTTCAACAATGATGTATCCAAACGTTCAAGAC 3552  
 QY 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 DB 3553 ACCTTCCGTGGGCAATTAATGGGCGCATCCAGCTTATGACCCGATGTCTCCAGCAC 3612  
 QY 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 DB 3613 CTCTTCTATCTCTTTGAGCGGTGCAATGAGCTGTGAGCTGTGTGTGTCAACAGACAGT 3672  
 QY 1080 GlyIleuLeuValLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSer 1099  
 DB 3673 GGGAGCTGCACTCAGCGGCAAGGATGACAATACCCCACTGTGTGCTCCCATGTGTG 3732  
 QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuAspValThrIle 1119  
 DB 3733 GTGATGTCAAGATGAGCTGTGACAGCGTGAAGGCGAGTGTGTGTGTGTGTGTGTGTGT 3792  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 DB 3793 ATCAAGAGAGTGTCTGCGCAACAGCGTGAACCGGCTTGAAGACATGTGTGAGAGAG 3852  
 QY 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
 DB 3853 CGCTTCTGTCAACCGCTGTGCGCGCTCTCTCAGAGGCGGTGTGTGTGTGTGTGTGTGT 3912  
 QY 1160 ProProAspHisValAlaValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
 DB 3913 CCGGTGAGAGAGCTTATATCTTCAACATCCAGAACACACAGACGTA---GGGGGACAC 3969  
 QY 1180 IleLeuAsnValSerLeuSerValGlyAlaProProGlyProGlyGlyPro----- 1197  
 DB 3970 GTGCTCAATGAGATTTCTCGGCGCTAGCTCCAGTGGGGCGGGGCGGTGTGACAGGG 4029  
 QY 1198 ProPheLeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThr 1217  
 DB 4030 CCTGTGCTCACTCCAGAGAGCTGTGAGAGGAGTGTGACGTGCGCGGGGCGGCGGTGCG 4089  
 QY 1218 AlaIleSerAlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCys 1237  
 DB 4090 GCTGCTCTCCCTGTGAGCTATCTGCTTCAACAACAAGTGTGCTGTGAGAGGCTGTGT 4149  
 QY 1238 GluAsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAla 1257  
 DB 4150 GAGAACTATCATGAATATGT 4209  
 QY 1258 SerSerSerValLeuAspArgProIleHisProValGlyIleuAspArgCysArgCysPro 1277  
 DB 4210 TGGGCTTCAAGCTGTGTGAGCCCAATCCAGCCCAATCGTGTGTGTGTGTGTGTGTGTGT 4269  
 QY 1278 ProGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCys 1297  
 DB 4270 CCGGATTCAGGGAGACTTTTGTGAGAGCGAGGTCAACCTCTGTCTACTCAACCAAGT 4329  
 QY 1298 GlyProHisGlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGly 1317

Db 4330 CGCAACGCGGAGCGCTGCGCGCGCGGAGGAGGCTACACGTCGCTGCCCGCGCGC 4389  
 QY 1318 TTTThrGlyValHisCysGluValIleSerAlaArgSerGlyValArgCysThrProGlyValAlaCys 1337  
 Db 4390 TTTCACCGAGAGAGAGCTGAGCTGAGACACGAGGCGCGCGCTGCGCGCGCGGCGTCTGC 4449  
 QY 1338 LysAsnGlyValThrCysValAsnLeuValGlyValPheIleCysAspCysProSer 1357  
 Db 4450 CGCAACGCGGCGCACCTGACCCGACCGCCCAACGCGCGCTTTCGCTGCGCGCGCA 4509  
 QY 1358 GlyAsp---PheGlyValProGlyCysGluValThrArgSerPheProAlaHisSer 1376  
 Db 4510 GGGGCGCGCTTCGAGGCGCGCGCTGCGAGGTGCTGCGCGCTTCCTCCGCGCGCAGTTTCG 4569  
 QY 1377 PheIleThrPheArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAla 1396  
 Db 4570 TTTCGTCATGTTTCCGCGCGCTGCGCGCGAGTTCCACTTCACGCTGCTCTGCTGCTGCG 4629  
 QY 1397 ThrLysGluArgAspGlyLeuLeuLeuTyraAsnGlyValArgPheAsnGlyValHisAspPhe 1416  
 Db 4630 ACAGTGCAGAGAGCGGCGCTGCTCTTCTACACGCGCGCGCTGAGACGAGAGCAAGACTTC 4689  
 QY 1417 ValAlaLeuGluValIleGluGluValGluLeuThrPheSerAlaGlyValSerThr 1436  
 Db 4690 CTGGCCCTGAGACTCGTGGCTGCGCGAGTGCCTGACATATTCACGCGGTGATTCAC 4749  
 QY 1437 ThrThrValSerProPheValProGlyValValSerAspGlyValGlnThrValGln 1456  
 Db 4750 ACCGTGTCAGCGCCCAACAGTTCCAGGGGCTTGAGTACCGGCGATGGCATACAGTGCAT 4809  
 QY 1457 LeuIleValTyraAsnLysProLeuLeuGlyValThrGlyLeuProGluGlyProSerGlu 1476  
 Db 4810 CTGAGATACCTACCAACAGCCCGGACAGATGCCCTAGGGGCTGACAGGCGCCCTCCAG 4869  
 QY 1477 GlnLysValAlaValValThrValAspGlyCysAspThrGlyValAlaAlaLeuArgPheGly 1496  
 Db 4870 GACAGGTGCTGCTGACAGCTGAGATGATGATGATGATGATGATGATGATGATGATGAT 4929  
 QY 1497 SerValLeuGlyValAsnTyraSerCysAlaAlaGlnGlyThrGlnGlyValSerLysSer 1516  
 Db 4930 GCTGAGATGGCACTACCTATGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4989  
 QY 1517 LeuAspLeuThrGlyProLeuLeuLeuGlyValValProAspLeuProGluSerPhePro 1536  
 Db 4990 CTGACCTGACGCGCGCTCTCTCTCTCTCTGAGAGTCTCCCACTCCCGGAGAACTTCC 5049  
 QY 1537 ValArgMetArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIle 1556  
 Db 5050 GTATCCCATAGGACTTCATCGCTGATGCGGAGCTTGACATGATGGCCGCGCAGATG 5109  
 QY 1557 AspMetAlaAspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysVal 1576  
 Db 5110 GACATGCGCGCTTTTGTGCGAATAATAGGACACATGCGAGGCTGCCAAGCAAGCTACAC 5169  
 QY 1577 ValCysAspSerAsnThrCysHisAsnGlyValThrCysValAsnGlnThrPheAlaPhe 1596  
 Db 5170 TTTTGTGACTCAGGCGCTGCGAGACAGTGGCTTCTGCTGCGAGCGCTGCGGAGGCTTC 5229  
 QY 1597 SerCysGluCysProLeuGlyPheGlyValLysSerCysAlaGlnGluMetAlaAsnPro 1616  
 Db 5230 AGCTGCACTGCGCTGCGGCTTTCGCGGCAACAGTGTCACTTACTTATGCGCCACATCC 5289  
 QY 1617 GlnHisPheLeuGlySerSerLeuValAlaThrHis---GlyLeuSerLeuProIleSer 1635  
 Db 5290 CACCATTTCCGTGGCAACGCGACCTGAGCTGAACTTTGGAGTGCATAGGCGTGTCT 5349  
 QY 1636 GlnProThrTyraLeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGln 1655  
 Db 5350 GTGCGCATGAGTACCTGGGCTGGGCTTTTCGACACGAGCNAACGAGGGGCTCTGATGCA 5409  
 QY 1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeu 1675

Db 5410 GTGCAAGCTGGGCAACAYNNACGCTCTTTCAGACTAGATCGGGGCTTACTGCTGTG 5469  
 QY 1676 SerVal---GluGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyValArgAla 1694  
 Db 5470 ACAGTACCAAGGAGGCTCGGAGC---CGTCTTCCATCTCTTCTGACCAAGGTACTGTC 5526  
 QY 1695 AsnAspGlyAspTrpHisIleAlaGlnLeuAlaLeu-----GlyAlaSerGlyValPro 1712  
 Db 5527 AGTATGCGCGGCTGACACATCTGCGGCTGAGTTGCAGAGAGAAACAGAGTGGCGCGCG 5586  
 QY 1713 GlyHisAlaIleLeu-----SerPheAspTyraGlnGlnArgAlaGluValAsnLeu 1730  
 Db 5587 GGCCACACATGCTTATAGTCTCACTGACCTTACCTTTCACAGACACCATGGCGAGT 5646  
 QY 1731 GlyProArgLeuHisGlyLeuHisLeuSerAsnIleThrValGlyValLe---ProGly 1749  
 Db 5647 GGGAGTGAAGCTCAGAGGCTTGAAGTAAAGAGCTCCAGTGGAGAGCTGCCCGCGCG 5706  
 QY 1750 ProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAsp 1769  
 Db 5707 AGTGCAGAGAGGCTCTCAGGAGTCTGAGTGTGCTCATTCAGAGGAGTGTGCTCGCTCC 5766  
 QY 1770 ThrProGluGlyValAsnSer---LeuAspProSerHisGlyGluSerIleAsnValGlu 1788  
 Db 5767 ACACCTCTGCTCCCGCGCTGCTACCCCGCACAC---CGAGTGAATCGGAG 5820  
 QY 1789 GlnGlyCysSerLeuProAspProCysAspSerAsnProCysProAlaAsnSerTyraCys 1808  
 Db 5821 CCGTGTGTGTGTGATCCAAACGCTGTGCTGTGGCGCTGCGCCACCTCAGCAGACACTGC 5880  
 QY 1809 SerAsnAspTrpAspSerTyraSerCysSerCysAspProGlyTyraTyraGlyValAspAsn 1828  
 Db 5881 CGGAGCTCTGGCAACCTTTCTTTCAGCTGCGAGCGCAGGTATCTACCGGCCAGGCTGT 5940  
 QY 1829 ThrAsnValCysAspLeuAsnProCysGlnHisGlnSerValCysThrArgLysProSer 1848  
 Db 5941 GTGATGCTGCTGCTCTTAACCTCTGTCAACAGAGATACATGCGGAGCTGCCAGCA 6000  
 QY 1849 AlaProHisGlyTyraThrCysGluCysProProAsnTyraLeuGlyProTyraCysGluThr 1868  
 Db 6001 GCCCCCAATGCTATACCTGTGATCTGTGAGGTGCTATTTGGGACACACTGTGAGCAC 6060  
 QY 1869 ArgIleAspGlnProCysPheArgGlyTyraTrpGlyValHisProThrCysGlyLysProAsn 1888  
 Db 6061 AGATATGACACAGAGTGCACAGGGGCTGTGTGGGAGCCCAACTGTGGCCCTGCGCAC 6120  
 QY 1889 CysAspValSerLysGlyPheAspProAspCysAsnLysThrSerGlyValLysCysCys 1908  
 Db 6121 TGTGATGTTCAAAAGGTTTGTATCCCACTGACACAGACAAATAGGCAAGTGTCTGC 6180  
 QY 1909 LysGluAsnHisTyraArgProProGlySerProThrCysLeuLeuCysAspCysTyraPro 1928  
 Db 6181 AAGAGATTCACATACGACCGCGGGGAGTACTCTTGCCTCCATGTGATCTACCTCT 6240  
 QY 1923 ThrGlySerLeuSerArgValCysAspProGluAspGlyValGlnCysPheValSerProGly 1948  
 Db 6241 GTGGGTTCACCTGCGCTGATGTCACCCCAACAGCGGAGATGCGCTGTGCGCCAGGA 6300  
 QY 1949 ValIleGlyArgGlyCysAspArgCysAspAsnProPheAlaGluValThrArgHisGly 1968  
 Db 6301 GCCCTTGGCGCGCAAGTGCACAGCTGTGACACTCCCTTGCAGAGTGCACACGCGCG 6360  
 QY 1969 CysGluValAsnTyraAspSerCysProArgAlaIleGluAlaGlyValLeuTrpProArg 1988  
 Db 6361 TCCCGGCTCTGATAGATGCTGCGCTTAAGTCTCGAGATGTGATGTGTGTGGCCCGC 6420  
 QY 1989 ThrArgPheGlyLeuProAlaValAlaAspProCysProLysGlySerPheGlyThrValVal 2008  
 Db 6421 ACAGATTTGGGTGCTCGGCCACAGTGCCTGTCTCCCGGGGAGCCCTGTGATCTGTG 6480  
 QY 2009 ArgHisCysAspGlnHisIleArgGlyTyraLeuProProAsnLeuPheAsnCysThrSerIle 2028  
 Db 6481 CGGCTGTGTATGAGGCGCAGGAGGTTGGCTGAGCGCGACCTTTCACTGTACTCTCCCT 6540

QY 2029 ThrPheSerGluLeuLysGlyPheAlaGluThrGluGlnArgGlnGluSerGlyLeuAsp 2048  
 Db 6541 GCCTTTGAGAGAGCTCAGTCTGCTGCTGAGGAGGCTTGAAGCTGAACAGAGAGGAGCTGAT 6600  
 QY 2049 SerGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAlaAlaThrGlnIleThrAlaGly 2068  
 Db 6601 ACCATGAGAGGCGCAAGAGAGCTGGCTCAGGAGGCTCAGGAGAGTGACTGGCGCACTGACAC 6660  
 QY 2069 TyrPheGlySerAspValLysValAlaTyrGlnLeuAlaThrArgLeuLeuAlaIleGlu 2088  
 Db 6661 TATTTTACCAGAAAGTTCGAGTCACTGCGCCGCTGGCGCTGGGCCCACTGCTGGCTTCGAG 6720  
 QY 2089 SerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGlnAsnLeu 2108  
 Db 6721 AGCATCAAGCAGGGGCTTCGGGTGACACACACAGAGATGCCACTTAAATGAAGAAATCG 6780  
 QY 2109 LeuArgValGlySerAlaLeuLeuAspThrAlaAsnLysArgHisIleProGluLeuIle 2127  
 Db 6781 CTGTGGGCGGGCTCGCACTGCTGTTGCCACAGACAGGGAGACTGTGGGGGGGCGCTGGGG 6840  
 QY 2128 GlnGlnThrGluGly-----GlyThrAlaTrieuLeuGlnHisTyrGluAlaTyr 2144  
 Db 6841 CAGGGGGGCGCTGGGGGGCTCCCAAGGCGCGGGGACTGTGTAAGGCACTGAGAGAGTAT 6900  
 QY 2145 AlaSerAlaLeuAlaGlnAsnMetArgHisIleThrTyrLeuSerProPheThrIleValThr 2164  
 Db 6901 GCAGCCACACTCGCAGAGAAATATGGAATCTCACTACCTGAATCCATAGGGGCTGGTAGCG 6960  
 QY 2165 ProAsnIleValIleSerValAlaArgLeuAspLysGlyAsnPhe-----AlaGlyAla 2182  
 Db 6961 CCTAATATCATGCTCAGCATTTGACCGGATGAGGACCCCAAGTTCCTCCCGGGGGGCGCGT 7020  
 QY 2183 LysLeuProArgTyrGluAla-----LeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
 Db 7021 GCCTAACCTCGCTACATGACAACCTCTTGGAGGCGAGATGCGCTGGGATCTTCACACC 7080  
 QY 2201 ThrValIleLeuProGluSerValPheArgGlnThrProProValValArgProAlaGly 2220  
 Db 7081 CATGTGCTGTGCTTCCCAAGTCCCAAGGCGGCAATCCCATGTGAATTTCTGCCACAAAGC 7140  
 QY 2221 ProGlyGlnAlaGlnLysProGluGlnLeuAlaArgArgGlnArgHisProGluLeu 2240  
 Db 7141 AGCAGCATAGAAAACCTCCACACCTCAAGTGTGGTCCCCCAGCAGCGCCGCGCAGAGGCA 7200  
 QY 2241 SerGlnGlyGlnAlaValAlaSerValIleIleTyrArgThrIleuAlaGlyLeuLeuPro 2260  
 Db 7201 GAGCCTGGGATCTCCATTATGATCTCTCGTTTACCGCACTTAAGGGGACTCTCCCT 7260  
 QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
 Db 7261 GCCCAGTTCAGGAGGAAGCGCGAGGTGCCAGGCTCTCTCAAGAACCCGCTCATGAATCC 7320  
 QY 2281 ProValValSerLysSerValHisAspAspGluLeuLeuProArgAlaLeuAspLys 2300  
 Db 7321 CCGGTGGTCAACGATGGCTGTGTCCACGAGCGCAACTTCTTAAGGGGAATCTTGAGATCC 7380  
 QY 2301 ProValThrValGlnPheArgLeuLeuGlnThrGlnAlaGlnTyrLysProIleCysVal 2320  
 Db 7381 CCCATCAGCCCTGAGATTTTCCTGTGACAGAGGAATCGAGAGCAGAGCGATCTGTGTG 7440  
 QY 2321 PheThrAsnHisSerLysLeuValSerGlyThrGlyGlyTyrPserAlaArgGlyCysGlu 2340  
 Db 7441 CAGTGGAGCCCACTGGCGCTGGCGGAGAGCATGTGTGTGTGACAGCAGCGAGATCGAG 7500  
 QY 2341 ValValPheArgLeuGlnSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
 Db 7501 CTGTGTGACAGGAATGGGTCCCAACCAAGTGTCTGCTCACCCGAGCAGGACTTGTGGG 7560  
 QY 2361 ValLeuMetAspValSerArgArgLys-----AsnGlyGlnIleLeuProLeuLysThr 2378  
 Db 7561 GTCTCTCATGATGCTCTTCCCTGTAGAGGCTGAGAGGGCGACCTGAGAGCTGTGACTGTG 7620

OY	2379	LeuThrTyValAlaIleuGlyValThreLeuAlaIleuLeuLeuThrPhePheLeu	2398
Db	7621	TTTACCACAGCGTGGCTGGCTGTCTGTGGCTGCGCTGGCTGACTGACGACATCTG	7680
OY	2399	ThrIleuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAla	2418
Db	7681	CTGAGCGCTGGCGAGCGCTCAAGTCCAAATGTGCGTGGATTCATGCAATATGTGGCAGCGCC	7740
OY	2419	LeuGlyIleuAlaGlnIleuValPheIleuGlyIleAsnGln	2432
Db	7741	CTGGGGGTGGCAGAGCTCTCTCTCTCTCTGGGATTCACAGAACCCCAATCAAGTGCAG	7800
OY	2432	-----	2432
Db	7801	GATCAGGCGCAGGGAACTTGTCCTGATGACCTACTGGCCGACGAGCGCTGGGGCCAA	7866
OY	2433	---AlaAspLeuProPheAlaCysGlyValIleAlaIleIleuLeuHisPheLeuTyIleu	2451
Db	7861	AATCCAGAGGTTCAGAGCTGGTGTGTGACATGACAGTGCATCTCCCTGCACTACTCTTCCTC	7920
OY	2452	CysThrPheSerThrAlaIleuLeuGlnAlaIleuHisIleuTyArgAlaIleuThrGlyVal	2471
Db	7921	AGCACTTTCGGTGGCTCTTCGTGACAGGGCTGTGACCTTACCGCAGTGAAGTTAAGCA	7980
OY	2472	ArgArgAlaAsnThrGlyProMetArgPheTyTyThrIleuGlyTyIleGlyValProAla	2493
Db	7981	CGCAACGTGACCGCGCGGCACATGCGTCTTACATGACATCCCTGGGCTGGGGCGTCCGTCT	8040
OY	2492	PheIleThrGlyIleuAlaValGlyIleuAspProGlnGlyTyIleGlyAsnProAspPheCys	2511
Db	8041	GTCGTGCTGGGCGCTGTGTGTGGGCTGTGACCCCTGAGGGCTATGAGAAACCTGACTTCTGC	8100
OY	2512	TyrIleuSerIleTyArgPheThrIleuIleTyPheSerPheAlaGlyProValAlaPheAlaVal	2531
Db	8101	TGATCTCAAGTCCACGAGCCCTCATCTGTGAGCGTTTCTGGCCCTGTATGTCTGGTCATA	8160
OY	2532	SerMetSerValPheLeuTyTyIleLeuAlaAlaArgAlaSerCysAlaAlaGlnArgGln	2551
Db	8161	GTCATGAAACGGACCAAGTTCCTCTGTCGCCGACATCTGCTCCACACAGGCGAAGG	8220
OY	2552	GlyPheGlyIleuTyIleGlyProValSerGlyIleuGlnProSerPheAlaIleuLeuLeu	2571
Db	8221	GAGGCCAAGAAAGACCTGTGACTACGAGACCTTTCGACGTCTCTTCGCTGCTCTGCTG	8280
OY	2572	LeuSerAlaThrTyIleuLeuAlaIleuLeuSerValAsnSerAspThrIleuLeuPheHis	2591
Db	8281	GTCAGTGCCTTCCTGGCTCTTGTGGCTCTCTGGACCTTCAACACAGCATCTTACCTTCCAC	8340
OY	2592	TyrIleuPheAlaThrCysAsnGlyIleGlnGlyProPheIlePheLeuSerTyValVal	2611
Db	8341	TACTCTCAGTGTGAGCTCTGCGGCGCTTCAGAGGCTGGCGGGTGTGCTGCTCTTGTGTCTC	8400
OY	2612	LeuSerTyIleuValAlaArgValAlaIleuTyIleuAlaCys---SerArgTyIleProSerPro	2630
Db	8401	CTAAATGCAGATGCTGGGGCTGCTGTATGCAGCGCTGTCTGGGACGAAAGGACGCGCT	8460
OY	2631	AspProAlaIleuThrTyIleSerThrIleuThrSerSerTyIleAsnTyIleProSerProTyIle	2650
Db	8461	GAGAGGCAAGGCCAGACACTGGGTGGGAGCTGGGGGCTTACACACACAGCGCTCTTCTT	8520
OY	2651	AlaAspGlyArgLeuTyIleGln---ProTyIleGlyAspSerAlaGlySerLeuHisSerThr	2669
Db	8521	GAGAGAGTGGCTCATCCGCGCATCTCTGGGCGGCTTCACCGCTCTCTCTGTAGACAGT	8580
OY	2670	SerArgSerGlyTyIleSerGlnPro-----SerTyIle-----	2680
Db	8581	GCCGCGCTCCGCGCGACCCAGAACACAGACAGCCAGCGGAGCGCGACACTCTCAGAGGAC	8640
OY	2681	ProPheIleuLeuTyIleGlnIleuSerAla-----LeuAsnProGly	2693
Db	8641	AATGTCTGTGTGACATGAGCTCAGCCGCTGACACACTGACACAGCTTCAGGCTCAT	8700
OY	2694	GlnGlyProProGlyIleuGly-----AspProGlySerLeuPheLeu	2707





Db 2473 TACTCAAGCTGGTACTAAGTCACTGACCGGTCCTTCATGATCACTGCTATGTGCAC 2532  
 Qy ValAsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisGlyThrVal 719  
 Db 2533 ATCAACATCAAGATGCGCACTCACTGCGGATCTTTCAAAATGCCCACTACTCATGTG 2592  
 Qy AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGlu 739  
 Db 2593 AGTGTGAATGAAGATCGGCGCAATGGGTAGCACAATAGTGTCATCAATGCTGCTGATGAT 2652  
 Qy AspThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2653 GACGTGGGTGAAATGCTCGTATCACTATCTCTGGAGGACCAACCGCCCGCTTCGCG 2712  
 Qy IleAspAlaAspThrGlyAlaValIleThrGlnAlaGluLeuAspTyrGlyAspGlnVal 779  
 Db 2713 ATTGATGCAAGCTAGAGAGCCATTACATACAGCCCATTAAGCTATGAGACCAAGTGTG 2772  
 Qy SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
 Db 2773 ACCTACACACCTGGCTATCAAGCTCGGGACAAATGGCATCCCAAGAGGACAGACTACT 2832  
 Qy TyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 2833 TATGTGAGGTGATGTCATGAATGAGTGAATGACAAATGCTCCACAAATTTGGCTCCAC 2892  
 Qy TyrGlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 2893 TATACAGGGCTGCTCTGAGAGATGCCCACTTTCACAGTGTCTGCGAATTCACAGCC 2952  
 Qy ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAsp 859  
 Db 2953 ACTGACCGGAGTGTCTCATGCGCAATGCGGGGTCAATGACACTTTCAGAAATGGTGAAGAT 3012  
 Qy GluAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgGluLeu 879  
 Db 3013 GGGGATGAGATTTACATGAGCCCACTCTGAAATGTGCTGTACAGTAAAGCGGCTA 3072  
 Qy AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyIleMetPro 899  
 Db 3073 GACCGGAGGAGTACAGTGTATGAGTGTGACGCTACAGAGTGTGACAGAGGTGTCCC 3132  
 Qy ProAlaArgThrProMetGluValIleValIleThrValLeuAspValAsnAspAsnPro 919  
 Db 3133 CCACTCGGAGCTCCAGTATCCAGTATCCAGTATGATGTGAGATGGAACGAAAGCACT 3192  
 Qy ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 3193 GCTTCCCAAGCTGAGAGTTTGAAGTGTGGGTGAAGAGATACCAATTTGGGCTCACTG 3252  
 Qy ValAlaArgValIleThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3253 GTGGCCAGATCACTGACGTGGACCCCTGACGAGAGGCCCAATGCGCATATATATACAG 3312  
 Qy IleValGluGlyAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3313 ATCTGTGAGGGGAATCTCCCTGAGCTGTTCAAATGACATCTTCTCTGAGAACTACAG 3372  
 Qy AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 3373 GCATCTATGACTAGACTATGAGGCTGCGCAAGAAATATGATTTGGTGCAGGCAACA 3432  
 Qy SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3433 TCTCTCTCTTGGTCAAGCGGGCACTGTGCACTGCGCTGCTGACAGAAATGACAAAC 3492  
 Qy ProProValIleuGlyAsnPheGluIleuPheAsnAsnTyrValIleAsnArgSerSer 1039  
 Db 3493 AGCCTGTGCTCAACACTTCAAGTCTTCAACAACATATGATCAACCGTTTCAGAC 3552  
 Qy SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAlaIleSerAspSer 1059  
 Db 3553 ACCTTCCGTGGGCAATTTGGGCGCATCCCACTTATGACCCCGATGTCTCCGACAC 3612

Qy LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1060  
 Db 3613 CTCTTTACTCTCTTGAAGGTGGCAATGAGCTGCACTGCTGTGATCAACAGCAACT 3672  
 Qy GluGluLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSer 1099  
 Db 3673 GGGAGCTGCACTGACCGGAAAGCTAGACATTAACCGCCACTGGTGGCTTCATGTTG 3732  
 Qy ValLeuValSerAspGlyValHisSerValIleThrAlaGlnCysAlaLeuArgValIle 1119  
 Db 3733 GTGACTGTCAAGATGAGCTGTGACAGCTGACAGGCGGCAATGTGTGTGCGGTGCATC 3792  
 Qy IleThrAspGluMetThrHisSerIleThrLeuArgGluLeuAspMetSerProGln 1139  
 Db 3793 ATCAACGAGAGATGTGTGCGCAACAGCTGACCGTGTGCTTGAGAAATGTGGCAGAG 3852  
 Qy ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrValIleThr 1159  
 Db 3853 CGCTTCTGTCAACCGCTGTGAGGCGCTGTGAGAGCGGTGGCTGCGCTGCTACG 3912  
 Qy ProProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
 Db 3913 CCGCTGAGAGCTCTTCACTTCAACATCCAAACAGACAGAGCTA--GAGGAGCAC 3969  
 Qy IleLeuAsnValSerLeuSerValGlyGlnProGlyProGlyGlyPro----- 1197  
 Db 3970 GTGCTCAATGTAGATTTCTCGCGCTAGCTCAACGTGGGCGGCGGCGGCTGACAGG 4029  
 Qy ProPheLeuProSerGluAspLeuGlnIleArgLeuTyrLeuAsnArgSerLeuLeuThr 1217  
 Db 4030 CCTGGTTCAGCTCCAGAGAGCTGCAAGAGAGTTGATGTGCGCGGCGGCGCTGAGG 4089  
 Qy AlaIleSerAlaGlnArgValLeuProPheAspAsnAlaCysLeuArgGluProCys 1237  
 Db 4090 GCTCGCTCCCTGCTGACGTAAGCTGCTTGCAGAGAGAGTGTGCTGAGAGAGCTGT 4149  
 Qy GluAsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAla 1257  
 Db 4150 GAGAACTATCATGAATAGCTGTCCGTCGCTCCGCTTGACTGTCGCGCTTCTGAGCC 4209  
 Qy SerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysPro 1277  
 Db 4210 TCGGCTTCAACCTGTTCGACCCATCCAGCCCATGCGCTGCGCTGCGCGCTGCGC 4269  
 Qy ProGlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCys 1297  
 Db 4270 CCGGATTCACGGAGACTTTTGGCAGACCGAGCTGCACTGTCTTCAACCACTGT 4329  
 Qy GlyProIleGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGly 1317  
 Db 4330 CGGACGGGAGAGCTGCGCGGCGGCGGAGGAGGCTACACGTGCGCTGCGCGCGCGC 4389  
 Qy TyrThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCys 1337  
 Db 4390 TTACCGGAGAGAGACTGCGAGCTGACACCGAGCGCGCGCTGTGCGCGCGCTGCG 4449  
 Qy LysAsnGlyTyrCysValAsnLeuLeuValGlyGlyPheLeuPheCysAspCysProSer 1357  
 Db 4450 CGGAGCGGGGCACTGACCGACCGGCGGCGGCGGCTTTCGCGTCCAGTGCAGGCGCA 4509  
 Qy GlyAsp--PheGluAspProTyrCysGlnValIleThrArgSerPheProAlaHisSer 1376  
 Db 4510 GCGGCGCTTGCAGAGCGCGCGCTGCGAGGTGTGCGCGCTTTCGCGCGCGCACTGCG 4569  
 Qy PheIleThrPheArgGlyLeuArgProGlnArgPheHisPheThrLeuLeuLeuSerPheAla 1396  
 Db 4570 TTGCTATGTTTGTGCGGCTGTGCGGCGAGCAATTCATACGCTGTGCTTCTGTTGCGG 4629  
 Qy ThrLeuGluArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIleAspPhe 1416  
 Db 4630 ACGGTGACAGAGAGCGGCTGCTTCTTACAAAGGCGCTGAGAGAGAGCAAGCACTTC 4689







Db 6790 GCGGCGCTGGGCGAGCGGGCCCTGGGGGCTCCCGACGACGCGGGAGCTGTGAGGAC 6849  
 Qy 2141 TGTGUAATATYALaseralaleuAlaGlnAsnMetArgH1serThyTyrLeuSerProPhe 2160  
 Db 6850 CTGGAGAGATATGACACCACTCGCAGAGAAATATGGAATCTCAATCTGAAATCCCAAG 6909  
 Qy 2161 ThrTleValThrProAniIleValIleSerValValArgLeuAspLysGlyAsnPhe--- 2179  
 Db 6910 GGGCTGTGACCGCTAATATCATGCTCAGCATTGACCGCATGAGACACCCAGTCTCC 6969  
 Qy 2180 ---AlaGlyAlaLysLeuProArgTyrGluAla-----LeuArgGlyGluGlnProPro 2196  
 Db 6970 CCGGGGCGCCGTGTACCTCGTACCATATGACCACTCTTGTAGAGCCAGAGTCCGTG 7029  
 Qy 2197 AspLeuGluThrThrValIleLeuProGluSerValPheArgGluThrProProVal 2216  
 Db 7030 GATTCACACCAACCATGTGCTGCTCTCCCAATCCCGACGCGCAATCCCATCTGAAATT 7089  
 Qy 2217 ArgProAlaGlyProGlyGluAlaGlnGluProGluGluLeuAlaArgArgLysArg 2236  
 Db 7090 CTGCCCCACAGCAGCAGCATAGAAACTCCACCACTCAAGTGTGTCCTCCCGACAGCC 7149  
 Qy 2237 HisProGluLeuSerGlnGlyAlaValAlaSerValIleTleTyrThrThrLeuAla 2256  
 Db 7150 CCGCAGAGCAGAGCCCTGGGATCTCATATCATCTCTCGTTTACCGCACCTTAGAG 7209  
 Qy 2257 GlyLeuLeuProH1AsnThyAspProAspLysArgSerLeuArgValProLysArgPro 2276  
 Db 7210 GGAATCTCTCTCCAGTTCCAGTCCAGGAGAAACCGCAGAGTCCAGGCTTCTCAAGACCC 7269  
 Qy 2277 IleIleAsnThrProValValSerIleSerValHisAspAspGluLeuLeuProArg 2296  
 Db 7270 GTCATNACTCCCGGTGTGTCAAGTGTGTTCACGAGCAGCACTTCCTAAGAGGA 7329  
 Qy 2297 AlaLeuAspLysProValThrValGlnPheArgLeuGluThrGluGluAlaGlyThrLys 2316  
 Db 7330 ATCTGAGAGCCCGCATCGAGCTTAGAGTTTGGCTGTACAGACAGAAATCGAGACAG 7389  
 Qy 2317 ProIleCysValPheThrAsnHisSerIleLeuValSerGlyThrGlyGlyTTPSerAla 2336  
 Db 7390 GCGATCTGTGTGAGTGGAGCCACCGTGGCTGGCGGAGCAGATGTGTGTGAGACGA 7449  
 Qy 2337 ArgGlyCysGluValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMet 2356  
 Db 7450 CCGGAGCTGCGAGCTGTGTGACAGAAATGGTCTCCACGACGATGTGTGTGACACCCGACA 7509  
 Qy 2357 ThrSerPheAlaValLeuMetAspValSerArgArgGlu-----AsnGlyGluIleLeu 2374  
 Db 7510 GGGACCTTTGGGCTCTCATGATGCTCTCCCGTGAAGAGCTGAGGGCGACCTGAG 7569  
 Qy 2375 ProLeuLysThrLeuThrTyrValAlaLeuGlyValThrLeuAlaIleLeuLeuThr 2394  
 Db 7570 CTGCTGGCTGTGTACCCACGATGATGATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 7629  
 Qy 2395 PhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsn 2414  
 Db 7630 GCAAGCATCTGTGAGCTGTGACCGCAGCTCAATGTCATGTCGTGGAGTCCATGCCAAT 7689  
 Qy 2415 LeuThrAlaLeuGluLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaArg 2434  
 Db 7690 GTGGACAGCCCGCTGGGGGTGGAGAGACTCTCTCTCTGGGAGATTTCACAGACCCAC 7749  
 Qy 2435 LeuProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPhe 2454  
 Db 7750 AATCAGCTGTGTGACATGTGATGTCATCTCTGTGACATCTTCTCTCAGACACTTC 7809  
 Qy 2455 SerThrAlaLeuLeuGluAlaLeuHisLeuTyrArgAlaLeuThrGluValAlaArgAspVal 2474  
 Db 7810 GCGGTGCTCTTGTGAGCGGGCTGCACTTACCGCATGCAAGTTGAGCCACACCAAGTGT 7869  
 Qy 2475 AsnThrGlyProMetArgPheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThr 2494  
 Db 7870 GACCGCGCGCCCATGCGCTTCTACCAATGCCCTGGGCTGGGGCGTCTGTGTGTGTGTGT 7929

Qy 2495 GlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysThrLeuSer 2514  
 Db 7930 GGCCTTGTGTGTGGCTGTGACCTGTAGGGCTATAGGAACCTGATCTTGTGTGTGTGTCA 7989  
 Qy 2515 IleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSerMetSer 2534  
 Db 7990 GTCCACGAGCCCTCATCTGAGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8049  
 Qy 2535 ValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla-----AlaGlnArgGlnLysPhe 2553  
 Db 8050 GGAACCATATTTCTCTGTGCTGCGCAGCATCTGTCTCAGAGGAGGAGGAGGCC--- 8106  
 Qy 2554 GlnLysGlyGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSer 2573  
 Db 8107 ---AGAAGACCTGTGACATGACCTTGTGAGCTCTCTCTGTGTGTGTGTGTGTGTGTGT 8163  
 Qy 2574 AlaThrTyrPheLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheIleTyrLeu 2593  
 Db 8164 GCTCTGTGCTCTTGTGGCTCTGTGACAGTCAACACACATCTTAGCCCTTCCACTACCTTC 8223  
 Qy 2594 PheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSer 2613  
 Db 8224 CATGCTGACCTCTGCGGCTCTCAGGGCTGTGGGCTGTGTGTGTGTGTGTGTGTGTGT 8283  
 Qy 2614 LysGluValArgLysAlaLeuLysLeuAlaCys---SerArgLysProSerProAspPro 2632  
 Db 8284 GCGATGTCTGTGGGCTGTGCTGTGATGCGACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 8343  
 Qy 2633 AlaLeuThrThrLysSerThrLeuThrSerTyrAsnCysProSerProTyrAlaAsp 2652  
 Db 8344 GGAAGCCAGCAGCAGCTGGGCTGTGAGCTGTGAGCTTCAACACACGCTCTTGTAGAG 8403  
 Qy 2653 GlyArgLeuTyrGln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArg 2671  
 Db 8404 AGTGGCTCATTCGATCATCTGTGGGCTGTGACCGCTCTCTGTGTGTGTGTGTGTGTGTGT 8463  
 Qy 2672 SerGlyLysSerGlnPro-----SerTyrIle-----ProPhe 2682  
 Db 8464 TCCGGCCGGAACCCAGACCAAGACAGCAGCGGGCCGACAGTCACTCAGGAGCAATGTC 8523  
 Qy 2683 LeuLeuArgGluGluSerAla-----LeuAsnProGlyGlnGly 2695  
 Db 8524 CTGTGTGACATGT 8583  
 Qy 2696 ProProGlyLeuGly-----AspProGlySerLeuPheLeuGluGly 2709  
 Db 8584 CCCACTGACCTGTGACGTGACATGTCATGCAATGCTGTGGCA----- 8628  
 Qy 2710 GlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGluAspAsp 2729  
 Db 8629 -----GACTCGACTGTGACATGACCTGTCTGTGAGAGAGAG 8667  
 Qy 2730 GlnSerGlySerTyrAlaSerThrHisSerSerAspSerGluGluGluGluGluGlu 2749  
 Db 8668 AGAAGTCTCTCATTCATCTTCAAGAAAGAGAGCAATGTGGCGGAGCGGGGGCGCGCTC 8727  
 Qy 2750 GlnGluGluAlaAlaPheProGlyGluGlnGlyTyrAspSerLeuLeuGlyProGlyAla 2769  
 Db 8728 CAACGGCACTGTGCGGAGGCCAG-----AGT 8757  
 Qy 2770 GlnArgLeuProLeuHisSerThrProLysAspGlyGlyProGlyLysAlaPro 2789  
 Db 8758 GAGAGGCTCTTCAACCCAC-----CCCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGT 8811  
 Qy 2790 TrpProGly-----AspPheGlyThr--- 2796  
 Db 8812 TGGCAGCCCTTGGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8871  
 Qy 2797 -----ThrAlaLysGluSerSerGlyAsnGlyAlaProGluGluArg 2810  
 Db 8872 AGCGCGCTGGGCTGTGACACCAAGAGATGACGTAAACAACACAGCAGACAGCCGCGC 8931

```

Qy 2811 LeuArgGluAsnGlyAspAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySer 2830
Db 8932 CTG---ACCAgTGGGAT-----GAGACTTCTCTG-----GCCCG 8964
Qy 2831 SerAlaGlnProHisLysGlyIleLeuLysLysCys---LeuProThrIleSerGlu 2849
Db 8965 GCCCAGCGCCAGAGAAAGGCACTCTAGAAACCGGTTGCCAATACCACTGGTCCACAG 9024
Qy 2850 LysSerSerLeuLeuArgLeuProLeuGluGlnCysThrGlySerSerArgGlySerSer 2869
Db 9025 ACCCGAGGTGCCCTCTGAG-----CTGTCCTGTGCCGTGCAGCCACCTTGGCCACCGT 9078
Qy 2870 Ala-----SerGluLysSerArgGlyGlyProProPro 2880
Db 9079 GCTGTGCCAGCTGCTCTTAAGTCGATCTATGCTGCGGGGACAGGGCAGCCTTTCA 9138
Qy 2881 ArgProProProArgGlnSerLeuGlnGluLeuAsn 2893
Db 9139 CAGCCAGCCAGCCGCTACTCTTCTAGAGAACGCTGAC 9177

```

Search completed: February 13, 2004, 20:34:19  
 Job time : 2604 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 19:39:01 ; Search time 305 Seconds

(without alignments)  
4230.042 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545  
Sequence: 1 MRSPTAGVPLPTPPPLLL.....AGTVDBDSGSEPLFNNFLH 2923

## Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xjh  
-O=/cgn2.1/USPTO.spool/US09916649/rnat.11022004.143817.20963/app.query.fasta\_1.3079  
-DB=Issued\_Patents\_NA -OPMT=fastest -SUFFIX=rm -MINMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09916649 @CGN 1.1 169 @rnat.11022004.143817.20963 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEJOBBY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database:

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1104	7.1	4566	2 US-08-465-976A-1	Sequence 1, Appl1
2	1104	7.1	4566	2 US-08-982-412-1	Sequence 1, Appl1
3	1032.5	6.6	4078	4 US-09-016-434-1066	Sequence 1066, Ap
4	881.5	5.7	4650	1 US-07-998-003A-102	Sequence 102, App
5	881.5	5.7	4650	1 US-08-453-274B-102	Sequence 102, App
6	881.5	5.7	4650	1 US-08-453-695A-102	Sequence 102, App
7	881.5	5.7	4650	1 US-08-268-161A-102	Sequence 102, App
8	881.5	5.7	4650	1 US-08-453-702A-102	Sequence 102, App
9	881.5	5.7	4650	2 US-09-039-639-102	Sequence 102, App
10	881.5	5.7	4650	3 PCT-US93-12588-102	Sequence 102, App
11	881.5	5.7	4650	5 PCT-US95-08071-102	Sequence 102, App
12	868.5	5.6	4104	1 US-07-998-003A-94	Sequence 94, Appl1

13	868.5	5.6	4104	1 US-08-453-274B-94	Sequence 94, Appl1
14	868.5	5.6 <td>4104</td> <td>1 US-08-453-695A-94</td> <td>Sequence 94, Appl1</td>	4104	1 US-08-453-695A-94	Sequence 94, Appl1
15	868.5	5.6 <td>4104</td> <td>1 US-08-268-161A-94</td> <td>Sequence 94, Appl1</td>	4104	1 US-08-268-161A-94	Sequence 94, Appl1
16	868.5	5.6 <td>4104</td> <td>2 US-08-453-702A-94</td> <td>Sequence 94, Appl1</td>	4104	2 US-08-453-702A-94	Sequence 94, Appl1
17	868.5	5.6 <td>4104</td> <td>3 PCT-US93-12588-94</td> <td>Sequence 94, Appl1</td>	4104	3 PCT-US93-12588-94	Sequence 94, Appl1
18	868.5	5.6 <td>4104</td> <td>5 PCT-US95-08071-94</td> <td>Sequence 94, Appl1</td>	4104	5 PCT-US95-08071-94	Sequence 94, Appl1
19	868.5	5.0	5610	4 US-09-262-537-19	Sequence 57, Appl1
20	780.5	4.9	5693	4 US-09-262-537-19	Sequence 19, Appl1
21	768.5	4.7	5491	4 US-09-262-537-5	Sequence 5, Appl1
22	733.5	4.6	4190	4 US-09-262-537-9	Sequence 9, Appl1
23	722	4.6	5587	4 US-09-262-537-1	Sequence 1, Appl1
24	722	4.6	5598	4 US-09-262-537-33	Sequence 1, Appl1
25	722	4.6	5598	4 US-09-262-537-33	Sequence 33, Appl1
26	697	4.5	2063	4 US-09-620-312D-1	Sequence 1, Appl1
27	691.5	4.4	5617	4 US-09-262-537-3	Sequence 3, Appl1
28	674	4.3	3072	4 US-09-620-312D-79	Sequence 79, Appl1
29	628.5	4.0	3033	1 US-08-453-695A-111	Sequence 111, App
30	628.5	4.0	3033	1 US-08-268-161A-111	Sequence 111, App
31	628.5	4.0	3033	2 US-08-453-702A-111	Sequence 111, App
32	628.5	4.0	3033	2 US-09-039-639-111	Sequence 111, App
33	628.5	4.0	3033	5 PCT-US95-08071-111	Sequence 111, App
34	627.5	4.0	4705	1 US-07-998-003A-96	Sequence 96, Appl1
35	627.5	4.0	4705	1 US-08-453-274B-96	Sequence 96, Appl1
36	627.5	4.0	4705	1 US-08-453-695A-96	Sequence 96, Appl1
37	627.5	4.0	4705	1 US-08-268-161A-96	Sequence 96, Appl1
38	627.5	4.0	4705	2 US-08-453-702A-96	Sequence 96, Appl1
39	627.5	4.0	4705	3 US-09-039-639-96	Sequence 96, Appl1
40	627.5	4.0	4705	5 PCT-US93-12588-96	Sequence 96, Appl1
41	627.5	4.0	4705	5 PCT-US95-08071-96	Sequence 96, Appl1
42	624.5	4.0	3353	1 US-08-453-695A-109	Sequence 109, App
43	624.5	4.0	3353	1 US-08-268-161A-109	Sequence 109, App
44	624.5	4.0	3353	2 US-08-453-702A-109	Sequence 109, App
45	624.5	4.0	3353	3 US-09-039-639-109	Sequence 109, App

## ALIGNMENTS

RESULT 1  
US-08-465-976A-1  
; Sequence 1, Application US/08465976A  
; Patent No. 5869632  
GENERAL INFORMATION:  
; APPLICANT: SOPPET, DANIEL R  
; APPLICANT: LI, YI  
; APPLICANT: ROSEN, CRAIG A  
; APPLICANT: RUBEN, STEVEN M  
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARBELL, BYRNE, BAIN GILFILLAN, CECCHI  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NJ  
; COUNTRY: US  
; ZIP: 07068  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,976A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY F  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-444  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 994-1744  
; TELEFAX: (201) 994-1744  
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 4566 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 212..2863  
 US-08-465-976A-1

Alignment Scores:  
 Pred. No.: 2,38e-60 Length: 4566  
 Score: 1104.00 Matches: 277  
 Percent Similarity: 52.40% Conservative: 116  
 Best Local Similarity: 36.93% Mismatches: 245  
 Query Match: 7.10% Indels: 112  
 Gaps: 22

US-09-916-849A-3 (1-2923) x US-08-465-976A-1 (1-4566)

QY 2203 ILeuPProgluSerValPheArgGlu-----ThrProValValArgProAla 2219  
 Db 26 GTCTGCCCCACAGCAGCAGCATAGAAACTCCACACCTCAAGTGTGCTCCCAACA 85  
 QY 2220 G1yProgluValaglIngluProgluIngluValaArgArgInlArgArgHleProglu 2239  
 Db 86 GCCCGG-----CGAGAG 97  
 QY 2240 LeuSerIngluValaValaAlaSerValIleIleTyArgThlLeuAlaGlyLeuLeu 2259  
 Db 98 CCAGAGCCTGGAGATCCATTATTAATTCCTCGTTTAAACCGACCTTAAGGGGAGCTCTC 157  
 QY 2260 ProHlaSerValPheProAspLysArgSerLeuArgValProLysArgProIleIleAsn 2279  
 Db 158 CTTCCCAAGTTCCAGGACAGACGCCAGAGTCCAGAGCTCTCCAGAACCCCGTATGAC 217  
 QY 2280 ThrProValValSerIleSerValHlaAspArgLugluLeuLeuProArgAlaLeuAsp 2299  
 Db 218 TCCCGGGGTGCAGCGTGGCTGTGTTCACAGGACCAACTCTTAAGGGGAATCTGGAG 277  
 QY 2300 LysProValThrValGlnPheArgLeuIngluThrgluValaGthThlPheProIleCys 2319  
 Db 278 TCCCCCAAGCTCAGAGATTTCCCTGTCTACAGACAGGAATCGAGACAGCGCATCTGT 337  
 QY 2320 ValPheTrpAsnHlaSerIleLeuValSerGlyThrglyGlyTyrPheSerAlaArgGlyCys 2339  
 Db 338 GTGCAAGTGGACCAACCTGGGCTGGCGAGACGATGTGTGTGACAGACAGCGGACTGC 397  
 QY 2340 GluValValPheArgAsnGluSerHleValSerCysGlnCysAsnHlaMetThrSerPhe 2359  
 Db 398 GAGCTGGTGACAGAGAAATGGGTCCCAACGACGATGCTGCACACCGGACAGGACCTTT 457  
 QY 2360 AlaValLeuMetLysValSerArgArgGlu-----AsnGlyGluIleLeuProLeuLys 2377  
 Db 458 GGGGCTCCTCAGAGAGCTCTCCCGGTGAGAGGCTGAGAGGCGACCTGAGACTCTGGCT 517  
 QY 2378 ThrLeuThyValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhe 2397  
 Db 518 GTGTTCACCAACGATGT 577  
 QY 2398 LeuThrLeuLeuArgGlyLeuLeuArgSerArgGlnHlaGlyIleArgArgAsnLeuThrAla 2417  
 Db 578 CTGTGAGCTGACGAGCTCAAGTCCAAATGTGGTGGATCCATGCGCAATGTGGAGCC 637  
 QY 2418 AlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPhe 2437  
 Db 638 GCCCTGGGGGGTGGACAGACTCTCTCTCTGTGGGAGATTCACAGACCCACAAACAGCTG 697  
 QY 2438 AlaCysThrValIleAlaIleLeuLeuHlaPheLeuThyIleCysThrPheSerTyrAla 2457  
 Db 698 GTGTGCACTGAGAGTGGCATCTCTGTGACACTACTTCTCTCAGACACTTGGCGGTGCTC 757

QY 2458 LeuLeuGluAlaLeuHlaSerValArgAlaLeuThrgluValArgAspValAsnThrgly 2477  
 Db 758 TTCTGACAGGGGTGACCTCTACCGCATGAGGTGAGCGACAGCGACCGCGGC 817  
 QY 2478 ProMetArgPheTyrTrpMetLeuGlyTyrGlyValProAlaPheIleThrglyLeuAla 2497  
 Db 818 GCCATCGCTTCTTACACATCCCTGGGCTGGGGGTGCTGTGTGTGTGTGTGTGTGTGT 877  
 QY 2498 ValGlyLeuAspProgluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyraAsp 2517  
 Db 878 GTGGGCTGAGCCCTAGGGCTTAGGAACTTCGATCTGCGATCTCAAGTCCAGAG 937  
 QY 2518 ThrLeuIleTrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeu 2537  
 Db 938 CCCCTCATCTGAGCTTGTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 997  
 QY 2538 TyrIleLeuAlaAlaArgAlaSerCysAla-----AlaGlnArgGlnIleGluLysVal 2556  
 Db 998 TTCTCTCTGCTGCTCCGACATCTCTCTCCACAGGCGAGAGGAGCGC-----AAGAA 1051  
 QY 2557 GlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTyr 2576  
 Db 1052 ACCTTGACACTGACCTTGCGAGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1111  
 QY 2577 LeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheHlaIleThr 2596  
 Db 1112 CTCTTTGGCTCTCTGCGACATCAACACAGACTCTGCTTCCATCTCCATCTCCATCTGGA 1171  
 QY 2597 CysAsnCysIleGlnIleProPheIlePheLeuSerTyValValLeuSerLysGluVal 2616  
 Db 1172 CTCTGCGGCTCCAGAGGCTGGCGGTGCTGTCTCTGTGTCTTAAATGCAAGTCT 1231  
 QY 2617 ArgLysAlaLeuLysValLeuAlaCys---SerArgLysProSerProAspProAlaLeuThr 2635  
 Db 1232 CGGGGCTGCTGATGCGACACTGTCTGGGCGAGAAAGCGCGCTGAGAGGAGAGGCCA 1291  
 QY 2636 ThrLysSerThrLeuThrSerSerTyraCysProSerProTyrAlaAspGlyArgLeu 2655  
 Db 1292 GCACCTGGGCTGGACCTGGGCTTCAACACACCGCTCTCTTGAAGAGAGTGGCTTC 1351  
 QY 2656 TyrGln---ProTyrGlyAspSerAlaGlySerLeuHlaSerThrSerArgSerGlyLys 2674  
 Db 1352 ATCCGATCACTGTGGGCGCTCCACCGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1411  
 QY 2675 SerGlnPro-----SerTyrIle-----ProPheLeuLeuArg 2685  
 Db 1412 ACCCAGACCAAGAGACGACGCGGCGCAGACTTACTAGGAGCAATGCTGTGTGCA 1471  
 QY 2686 GluGluSerAla-----LeuAsnProGlyGlnIleProProGly 2698  
 Db 1472 CATGGCTCAGCGCTGACCACTGACCAAGCTTCCAGGCTCATGTGGCTCCACTGAC 1531  
 QY 2699 LeuGly-----AspProGlySerLeuPheLeuGluGlnArgGln 2712  
 Db 1532 CTGAGCTGACATGTTCATGCAATGCTGGCGCA----- 1567  
 QY 2713 GlnHlaAspProAspThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerGly 2732  
 Db 1568 -----GACTCCCACTCTGACATGACTCTCTCTTGGAGAGAGAGAGACTCTC 1615  
 QY 2733 SerTyrAlaSerThrHlaSerSerAspSerArgGluGluGluGluGluGluGluGlu 2752  
 Db 1616 TCATTCATCTTCAAGAAAGAGAGACATGCGCGGACGCGGGGCGCTTCCACAGCCA 1675  
 QY 2753 AlaAlaPheProGlyGluGluGlnIleTyrAspSerLeuLeuGlyProGlyAlaGluArgLeu 2772  
 Db 1676 CTGTGCGAGAGCGCAG-----AGTGAAGAGCTC 1705  
 QY 2773 ProLeuHlaSerThrProLysAspGlyGlyProGlyProGlyLysValaProTyrProGly 2792  
 Db 1706 CTCACCCAC-----CCCAAGATGTGATGAGCAATGACTCTGTGCTTATGTGGCAGGC 1759  
 QY 2793 -----AspPheGlyThr----- 2796

```

Db 1760 CTGGGGAGTGGCGAGGCCCTGTGCTCTGCAAGCTTGGGGCTCTGAAAGCGCTG 1819
Qy 2797 -----ThralatylglserserlyAsnlyalProglugluArgleuArglu 2813
Db 1820 GGGGTGACACACAGAGATGACGCTACACACACAGCCAGCCCGGCTG--ACC 1876
Qy 2814 AsnlyAspAlaIleuSerArggluglySerleuIleuProgluIleuSerAla 2833
Db 1877 AGTGGGGAT-----GAGACTCTCTG-----GGCCGGGCCCAAGCC 1912
Qy 2834 ProHialysgllyleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2852
Db 1913 CAGAGAAAGGACCTCGAAGAACCGGTGCATATACCTGCTGACACAGCCAGAGT 1972
Qy 2853 LeuIleuArgleuProleuIleuIleuIleuIleuIleuIleuIleuIleu 2870
Db 1973 GCGCCCTGAG-----CTGCTCTGCTGCGCGAGCCACCTTGGCCACCGTCTG 2026
Qy 2871 -----SerIleuIleuIleuIleuIleuIleuIleuIleuIleu 2883
Db 2027 GCTGCTCTTACCGTCCATCTATGCTGGCGGGGACAGGAGCTTTCAAGCCAGCC 2086
Qy 2884 ProArgIleuSerleuIleuIleuIleuIleuIleuIleuIleuIleu 2893
Db 2087 AGCGGCTACTCTCTAGAGAACAGCTGGAC 2116

RESULT 2
US-08-982-412-1
; Sequence 1, Application US/08982412
; Patent No. 5958729
; GENERAL INFORMATION:
; APPLICANT: SOPEET, DANIEL R
; APPLICANT: LI, YI
; APPLICANT: ROSEN, CRAIG A
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE,
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,412
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF181PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; NAME/KEY: CDS
; LOCATION: 212..2863
; US-08-982-412-1

```

```

Alignment Scores:
Pred. No.: 2,386-60 Length: 4566
Score: 1104.00 Matches: 277
Percent Similarity: 52.40% Conservative: 116
Best Local Similarity: 36.93% Mismatches: 245
Query Match: 7.10% Indels: 112
DB: 2 Gaps: 22

US-09-916-849A-3 (1-2923) x US-08-982-412-1 (1-4566)
Qy 2203 IleleuProgluSerAlaPheArgIleu-----ThProProValAlaArgProAla 2219
Db 26 GTCTCCCAACAGAGACAGACATGAAATCCACCACTCAAGTGTGTCCCAACA 85
Qy 2220 GlyProgluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2239
Db 86 GCGCCG-----CCAGAG 97
Qy 2240 LeuSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2259
Db 98 CCAAGCCCTGGATCTCCATTAATTCCTCTGTTACCGACCTTAGGGGAGCTGCTC 157
Qy 2260 ProHialysIleuProArgIleuIleuIleuIleuIleuIleuIleuIleu 2279
Db 158 CCGCCAGCTTCAGGACAGACCGCAGGTGCAGGCTTCTCAGAACCCGTCATGAAC 217
Qy 2280 ThrProValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2299
Db 218 TCCCGGTGTGACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 277
Qy 2300 LysProValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2319
Db 278 TCCCGCATGAGCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 337
Qy 2320 ValPheIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2339
Db 338 GTGACAGTGGACCCACCTGCTGGGAGACAGCATGTGTGTGTGTGTGTGTGTGT 397
Qy 2340 GluValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2359
Db 398 GAGCTGTGACAGAGATGGTCCACGAGGTGTGTGTGTGTGTGTGTGTGTGTGT 457
Qy 2360 AlaValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2377
Db 458 GGGGTCTCATGATGCTCTCCCGTGAAGCTGAGGGGACCTGAGCTGGCTGCT 517
Qy 2378 ThrIleuThrValAlaIleuIleuIleuIleuIleuIleuIleuIleu 2397
Db 518 GTTTCACCCACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 577
Qy 2398 LeuThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2417
Db 578 CTGCTGAGCTGCGAGCTGCAATGCAATGCGGTGATCCAGGCCAATGGAGGCC 637
Qy 2418 AlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2437
Db 638 GCGCTGGGGGTGACAGCTCTCTCTGCTGGGATTCACAGACCCACATCAGCTG 697
Qy 2438 AlaCysThrValIleuIleuIleuIleuIleuIleuIleuIleuIleu 2457
Db 698 GTGTGACATGACAGTGCATCTCTGACATCTCTCTGACACCTTCCGCGTGC 757
Qy 2458 LeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2477
Db 758 TTGCTGAGAGGGGTGACCTTACCGCATGACGTTGAGCCAGCAGCAGTGAAGCG 817
Qy 2478 PrometIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2497
Db 818 GGCATGCGCTTACATGACCTGAGGCTGGGCGTCCCTGCTGTGTGTGTGTGTGT 877
Qy 2498 ValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2517
Db 878 GTGGGCTGAGACCTTGAAGGGCTATGGGAACCTGATCTGTGTGTGTGTGTGTGT 937

```

```

Qy 2518 ThrLeuIleTPSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeu 2537
Db 938 CCCCTCATCTGAGCTTGTGTGGCCCTGTGTGCTGTGATGATGATGAAGGAGCCATG 997
Qy 2538 TyrIleLeuAlaAlaArgAlaSerCysAla---AlaGlnArgGlnGlyPheGlnLys 2556
Db 998 TTTCTCTGCTGCTGCGGACATCTGCTGCACAGGAGGAGGAGGAGCC---AAGAG 1051
Qy 2557 GlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrp 2576
Db 1052 ACCCTGACATGACCTTCCTGAGCTCTTCTGCTGCTTCTGCTGCTGCTGCTGCTG 1111
Qy 2577 LeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheLeuPheAlaThr 2596
Db 1112 CTCTTGGGCTCTCTGAGCTGACACACAGCATCTTCACTTCACTTCACTTCACTTCA 1171
Qy 2597 CysAsnGlyIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerLysVal 2616
Db 1172 CTCTGCGGCTCTGACGAGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1231
Qy 2617 ArgLysAlaLeuLysLeuAlaCys---SerArgLysProSerProAspProAlaLeuThr 2635
Db 1232 CGGCTGCTGCTGATGCCAGCTGTCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1291
Qy 2636 ThrLysSerThrLeuThrSerSerTyrAsnGlyProSerProTyrAlaAspGlyArgLeu 2655
Db 1292 GCACCTGGGCTGGGAGCTGGGCTGACCAACACAGGCTCTTGTGAGAGAGAGTGGCTC 1351
Qy 2656 TyrGln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerLys 2674
Db 1352 ATCCGATCATCTGAGGCGCTGCACCGCTCTCTGTGAGAGAGGCGGCTGCGGCGG 1411
Qy 2675 SerGlnPro---SerTyrIle---ProPheLeuLeuArg 2685
Db 1412 ACCGAGACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1471
Qy 2686 GlnGlnSerAla---LeuAsnProGlyGlnGlyProGly 2698
Db 1472 CATGCTGACGCGCTGACACATGACACAGCTCCAGGCTCATGCGGCCCATGAC 1531
Qy 2699 LeuGly---AspProGlySerLeuPheLeuGlnGlyGlnAspGln 2712
Db 1532 CTGACGCTGACATTTCCATGAGATGCTGCGCA--- 1567
Qy 2713 GlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAspGlnSerGly 2732
Db 1568 ---GACTCGAAGCTGACAGTACCTGTCTTGTGAGAGAGAGAGAGAGTCTC 1615
Qy 2733 SerTyrAlaSerThrHisSerSerAspSerGlnGlnGlnGlnGlnGlnGlnGln 2752
Db 1616 TCCATTTCCATCTTCAAGAAAGCAGGACATGCGCGGAGCGCGGCTTCCACAGGCCA 1675
Qy 2753 AlaAlaPheProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2772
Db 1676 CTCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1705
Qy 2773 ProLeuHisSerThrProLysAspGlyGlyProGlyProGlyAlaProTrpProGly 2792
Db 1706 CTCACCCAC---CCCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1759
Qy 2793 --- 2796
Db 1760 CTGAGGAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1819
Qy 2797 ---ThrAlaGlyGlnSerSerGlyAsnGlyAlaProGlnGlnGlnGlnGlnGln 2813
Db 1820 GGGCTGAGACACAGGAGATGACATCAACACAGGAGGAGGAGGAGGAGGAGGAGGAG 1876
Qy 2814 AsnGlyAspAlaLeuSerArgGlnGlySerLeuGlyProLeuProGlySerSerAlaGln 2833
Db 1877 AGTGGGAGT---GAGACTTCTCTG---GACCGGAGGAGGAGGAGGAGGAGGAGGAG 1912

```

```

Qy 2834 ProHisGlyIleLeuLysLysCys---LeuProThrIleSerGlnLysSerSer 2852
Db 1913 CAGAGAAAGGAGCTCTGAAAGAACCGGTTGCAATACCACTGGTGGCAGACCCGAGT 1972
Qy 2853 LeuLeuArgLeuProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2870
Db 1973 GCCCGTAG---CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2026
Qy 2871 ---SerGlnLysSerArgGlyGlyProProArgProPro 2883
Db 2027 GCTGCTTACGCTGACATGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2086
Qy 2884 ProArgGlnSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2933
Db 2087 AGCGCTACTCTTCTAGAGAAACAGCTGAC 2116

```

## RESULT 3

```

US-09-016-434-1066
/ Sequence 1066, Application US/09016434
/ Patent No. 6500938
/ GENERAL INFORMATION:
/ APPLICANT: Janice Au-Young
/ APPLICANT: Jeffrey J. Sellhammer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
/ TITLE OF INVENTION: PATHWAY GENE EXPRESSION
/ NUMBER OF SEQUENCES: 1490
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/016,434
/ FILING DATE: HEREMITH
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Zeller, Karen J
/ REGISTRATION NUMBER: 37,071
/ REFERENCE/DOCKET NUMBER: PA-0002 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 1066:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4078 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GENBANK
/ CLONE: g1107686
/ US-09-016-434-1066

```

## Alignment Scores:

```

Pred. No.: 6.96e-56 Length: 4078
Score: 1032.50 Matches: 280
Percent Similarity: 47.85% Conservative: 132
Best Local Similarity: 32.52% Mismatches: 348
Query Match: 6.64% Indels: 101
DB: 4 Gaps: 23

```

US-09-916-849a-3 (1-2923) x US-09-016-434-1066 (1-4078)

QY 179 ProGlnPheGlnProSerTYrGlnAlaThrValProGluuGlnProAlaGlyThr 198  
DB 1567 CCTGAATTTCCAGACAGCGGTCAAGCTGTTTGCATGAGAACGCGCATTTGGACT 1626  
QY 199 ProValAlaSerLeuArgAlaIleAspProAsnGluGlyValAlaGlyLeuGly 218  
DB 1627 ACTATCGTAGCGCTGAGCTGATGACCTGATGAGGAGTGAAGATGGTACGTAAC 1686  
QY 219 ThrMetAspAlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThr 238  
DB 1687 AGATGCGCAATTTA-----AATCATGTGCGCTTGGCATGACCATTTCACT 1734  
QY 239 GlyAlaValThrThrAlaGluGluLeuAspArgGlu---ThrIysSerThrIleValPhe 257  
DB 1735 GGTGCGGTGAGTACGTACAGAAACCTGCACTACGAACTGATGCGGTGTTTATCTCG 1794  
QY 258 ArgValThrAlaGlnAspHisGlyMetPro---ArgArgSerAlaLeuAlaThrLeuThr 276  
DB 1795 AGGATTCGTGCATCAGACTGGGGCTTGGCTACCGCGGGAAGTCCGAAGCTCTTGCTACA 1854  
QY 277 IleLeuValThrAspThrAsnAspHisAspProValPheGluGlnGlnGlyIleValPhe 296  
DB 1855 ATTACTCTCAATTAATCTGAATGACACACACTTTGTTGAGAAATAATTTGTGAAGG 1914  
QY 297 SerLeuArgGluLeuLeuGluValGlyTYrGluValLeuThrValAlaThrAspGly 316  
DB 1915 ACAATTCGCCAGATCTAGCGGTGGAGCAATAACCACTCTTCTGCTATGTATGCA 1974  
QY 317 AspAlaProProAsnAlaAsnIleLeuTYrArgLeuLeuGluGlySerGlyIleSerPro 336  
DB 1975 GAT-----GACTTCAGTTGGTACAGTACATGATTA-----GCTGAAATTA 2019  
QY 337 SerGluValPheGluIleAspProArgSerGlyValIleArg---ThrArgGlyProVal 355  
DB 2020 CTGATCTTGTGTTAATCAACCCCACTCGGGGTATGTGCATTAAGGATCGCTAAG 2079  
QY 356 Asp-----ArgGluGluValGluSerTYrGlnLeuThrValGluAlaSerAspGlnGly 373  
DB 2080 GATGGCTTAGGTCGAAGGTGTCTTCCACAGCTGTGAGATCAACACTACAGATGAGAA 2139  
QY 374 ArgAspProGlyPro-----ArgSerThrAlaIle----- 384  
DB 2140 AATTTGCCACACCATTAATATACATACATACAGTGGCTCCAGTCAACAAGCTGTAAAC 2199  
QY 384 ----- 384  
DB 2200 TTGCAGTGTGAGAGACTGGTGTGGCAAAATGCTGGCAGAGAAAGCTCCGACGCAAT 2259  
QY 385 -----ValPheLeuSerValGluLeuAspAsnAsp 394  
DB 2260 AATTAACAACAAGGAGAGTGAAGATATTTCTTCATTTCTCACTCTGCAATGCT 2319  
QY 395 AsnAlaProGlnPhe---SerGluTYrArgTYrValIleGlnValArgIleAspValThr 413  
DB 2320 CACATACCGCAGTTTAGAGCACTTCCAGCTGATTCAGGTATTAAGGAAACCAAGCT 2379  
QY 414 ProGluAlaProValLeuArgValThrAlaSerAspArgAspIleSerAsnAlaVal 433  
DB 2380 GTGGGTTCAGTGTAAATTTTCATGAACTCACTGACCTGACACTGGCTTCATGAAA 2439  
QY 434 ValHisTYrSerIleMetSerGlyAsnAlaArgGlyGlnPheTYrLeuAspAlaGlnThr 453  
DB 2440 CTGTGCTATGCTTTCTGAGGAATGAGATAGTTGCTTCATGATGATGAGAAACA 2499  
QY 454 GlyAlaLeuAspValValSerProLeuAspTYrGluThrIleGlyIleValArg 473  
DB 2500 GGAATGTGAATAATTTATCTCTCTTGAACCGTGAACCAACAACAATACCTCGAAT 2559  
QY 474 ValArgAlaGlnAspGlyIleArgProProLeuSerAsnValSerGlyLeuValThrVal 493  
DB 2560 ATTAACGCTATGACCTTGGGATACCCCAAGAGCT---GCGTGGCGTCTTCTACATGTC 2616

QY 494 GlnValLeuAspIleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThr 513  
DB 2617 GGTGTGTGATGCCAATGATATATCCACCGAGTTTACAGAGAGCTAATTTTGGAA 2676  
QY 514 ValLeuGlnSerValProLeuGlyTYrLeuValIleuHisValGlnAlaIleAspAlaAsp 533  
DB 2677 GTGAGTGAAGACAGAGGATCATATGTAATATCATCCAGGTGAAGCCACAGATTAAGC 2736  
QY 534 AlaGlyAspAsnAlaArgLeuGluTYrArgLeuAlaGlyIleHisAspPheProPhe 553  
DB 2737 CTGGGCGCCACAGCACGTGACGTATCTCAAT-----CTTACAGACACACACATTT 2790  
QY 554 ThrIleAsnAsnGlyThrGlyTYrPheSerValAlaAlaGluLeuAspArgGluVal 573  
DB 2791 TCAATTAACAGCGTACCGGCTGTGTAAATCATCCGACCGCTCTGATGAGAGCTGAC 2850  
QY 574 AspPheTYrSerPheGlyValGluAlaArgAspHisGly-----ThrProAlaLeuThr 591  
DB 2851 CATGACACTCTTAAGATGAGCCAGGAGCAAGCCAGAGAAAGAGCTCAGCTGTC 2910  
QY 592 AlaSerAlaSerValSerValThrValLeuAspValAsnAspAsnProThrPheThr 611  
DB 2911 TCCACTGTGCTGTGAAGTACACTAGAAAGATGTATATGACACACCACTTATAT 2970  
QY 612 GlnProGluTYrThrValArgLeuLeuGluAspAlaAlaValGlyThrSerValValThr 631  
DB 2971 CCACCTAATTAATCGTGTGAAGATCCAGAGATCTTCCAGAAAGAACCTCATCATGTG 3030  
QY 632 ValSerAlaValAspArgAspAlaHisSerValIleThrTYrGlnIleThrSerGlyAsn 651  
DB 3031 TTAGAAGCCACGATCCGTATTAAGTCA-----TCTGATCAG 3069  
QY 652 ThrArgAsnArgPheSerIleThrSerGlnSerGlyIle----- 664  
DB 3070 GAG-----AGATACAGCTTGTGACCAAGAGAAAGAACTTGATGTGATAACTC 3123  
QY 665 ---GlyLeuValSerLeuAlaLeuProLeuAspTYrIleGluArgGlnTYrValLeu 683  
DB 3124 AGTGAAGCAGTTAGATGTGTCCAGCAGTGTGACCTTGAAGAACCAAGTGTATATCTC 3183  
QY 684 AlaValThrAlaSerAspGlyThrArg-----GlnAspThrAlaGlnIleVal 699  
DB 3184 ACTGTGAGGCGCAAGACAGAGAAAGCAAGTTCTCTGTCTTCACTTGTGATGTGA 3243  
QY 700 ValAsnValThrAspAla-----AsnThrHisArgProValPheGlnSerSerHisTYr 717  
DB 3244 GTTGAAGTGTGATGTGAATGAGAACTGCACCAACCGTGTTCAGCTTGTGGAA 3303  
QY 718 ThrValAsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThr 737  
DB 3304 AAGGGAACAGTGAAGAAAGATGCACCTGTGTTCATTTGATATGACGCTGCGCTCAT 3363  
QY 738 AspGluAspThrGlyGluAsnAlaArgIleThrTYrPheMetGluAsp-----SerIle 755  
DB 3364 GATGAGAGCGCGGAGAGATGGGAGATCCGATCTCATTAGAGATGCTGCGCTT 3423  
QY 756 ProGlnPheArgIleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTYr 775  
DB 3424 GGTGTTTCAAAATAGTGAAGACAGGTGTCATAGAGACGTGCAGATCGACGTGACCGT 3483  
QY 776 GlnAspGlnValSerTYrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIle 795  
DB 3484 GATGACACTCCCATTAATTTGGCTAACATCTTTGCAACCGAAGAGGCTGCTGCTCTT 3543  
QY 796 SerAspThrThrTYrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPhe 815  
DB 3544 TCATGCTCATAGATCATCATAGAGTGAAGATGATCAATGACAAATGCAACCAAGACA 3603  
QY 816 LeuArgAspSerTYrGlnGlySerValTYrGluAspValProProPheThrSerValLeu 835  
DB 3604 TCAGAGCGCTGTTTATTAACCAAGAAATCATGAAATTCCTTAAGATGATCTGTGTC 3663  
QY 836 GlnIleSerAlaThrAspArgAspSerGlyLeuAsnGlyArgValPheTYrThrPheGln 855



```

Db      3664 CAGATTCAGGCGATTGATTCAGATTGCTCTATATGACAGAGCTCATGATCAAAATTACA 3723
Qy      856 G1YGIYAspArgLYAspArgLYAspPhe1IleValG1uSerThrSerG1YIleValArgThr 875
Db      3724 AGTGAATCA---CAAGATTCCTTTCAATATCAATCTTAATAAGATCTCATCAACT 3780
Qy      876 Leu---ArgArgLeuAspArgLYAsnValAlaGlnTyrValLeuArgAlaTyrAlaVal 894
Db      3781 ACGTCAAGAACTAGACCGGAGAACAGACAGATGACACATATATAGAGTTACTGTGACA 3840
Qy      895 AspLYAspG1YMetProProAlaArgThrProMetGluValThrValThrValLeuAspVal 914
Db      3841 GACATGATGATGATCCCAAAATCAACATTCAGAAAGATTCATTTGAAATCTTGATGAA 3900
Qy      915 AsnAspAsnProProValPheGluGlnAspGluPheAspValPheValGluGlu----- 932
Db      3901 AATACAAACAACTCATGATTTCTGCAAAAGTTCTACAAATCATCACTCCCTGAGCGGAA 3960
Qy      933 -----AsnSerProIleGlyLeuAlaValAlaArgVal 943
Db      3961 AAGCCAGACCGAAGAAATGACAGACGAGGAGCGCTC-----TATCGCGTC 4008
Qy      944 ThrAlaThrAspProAspArgLYThrAsnAlaGlnIleMetTyrGlnIleValGluGly 963
Db      4009 ATAGCCACCGACAGAGATGAGGCGCCCAATGCAAAATCTTCACAGCATGAGAGAGCGG 4068
Qy      964 Asn 964
Db      4069 AAT 4071

RESULT 4
US-07-998-003A-102
; Sequence 102, Application US/07998003A
; Patent No. 5643781
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; ADDRESSEE: Bicknell
; STREET: 20 South Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/998, 003A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5643781and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 30903
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: 495..4103
US-07-998-003A-102
Alignment Scores:
Prod. No.: 3 25e-46 Length: 4650
Score: 881.50 Matches: 343
Percent Similarity: 38.23% Conservative: 146
Best Local Similarity: 26.82% Mismatches: 452
Query Match: 5.67% Indels: 339
DB: 1 Gaps: 46

US-09-916-849a-3 (1-2923) x US-07-998-003A-102 (1-4650)
Qy      3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeu 22
Db      513 AGCCGAGCGCCCTGGGGGAGACAGCGCTACTGCTGCTCCATGCTGTATACATGCTGCTC 572
Qy      23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCyArgSerLeuGlySer 42
Db      573 CTGCTGGCTCATCCCAAG-----CCAGCCACTGGGTAGTGA 613
Qy      43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuCyAspSerSerAla 62
Db      614 CAAGGTCCGGAGAGAACACCAACCACTT-----CATGG-----GAGCTGCG 661
Qy      63 SerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82
Db      662 AGCCGACTATGTTT-----TCCAGATGTGGGGCACTGTACAAAGTACAGAGTGGGTGC 715
Qy      83 LeuValPro-----HisHisAspGlyLeuArgVal 92
Db      716 CCCGTACTTCGGGTGATGACAGACAGGTACATTTTACACCAACGACCTC----- 769
Qy      93 TrpCyProGluSerGlnAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
Db      770 -----CATGACCGTAGAGGGGCTCCGTAATGACAGAACAGCT 808
Qy      111 ProTyrSerCysArgLeuLeuGlyTyleGlyHisLeuSerProGlnGlyLeuThr 130
Db      809 CCTGTGTATCCCTGATCCCTGATGAGTTGAGGT----- 841
Qy      131 LeuProGluGlnHisProCysLeuYsaIaPArgSerLeuArgCysGlnSerCysIysLeu 150
Db      842 -----ATCTATCACACACTCTGTGCAAGATC----- 868
Qy      151 AlaGlnAlaProGlyLeuArgAlaGlyIuArgSerProGluGlnSerLeuGlyArg 170
Db      869 ---GAGCCCCCGGCTGTAGAGGGCCAGATAGAGT----- 901
Qy      171 ArgLYAsArg-AsnValAsnThrAlaProGlnPheGlnProProSerTyrGlnAlaThrVal 190
Db      902 ACAAGACATCAATGACACACACA-----CCCAACTTCGCTCACCAGTCACTCTGGCCAT 958
Qy      190 lProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206
Db      959 CCTTGAAACACCAACATCGGCTCACTTCCCACTCCGCTGCTCA----- 1007
Qy      206 eaAspProAspGluGlyAlaGlyArgLeuGluTyrThrMetAspAlaLeuPheAspSe 226
Db      1008 -GACGTGATGTGCTGCTCCCAAGGAGTGGCATTCATGAGCTGACAGTGGCAGAGACCA 1066
Qy      226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGluG 246
Db      1067 GAGAGAGAACAA-----CCACAGCTCATTTGATGGGCAA 1102
Qy      246 uLeuAspArgGluThrLYSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266
Db      1103 CTGGAACCTGAGCGGTGGGAGCTCTTATGACCTCAACCACTCAAGTGCAGATGGCGGCG 1162
Qy      266 tProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286
Db      1163 CCCCCACGCGCCACAGATGCGCTGCTGCTGATCACCGTGTGACACCAATGACAAACGC 1222

```



```

Db 2997 -----GCCCTGGCGTTCTTGCGCTACTGACAGACGAGGAGGCCAAAGCTGTGA 3049
Qy 983 -----AapLeuAspTyrGluAspArgProGluTyrValle 994
Db 3050 CCAGGCTGTGAAGAGAGACCAAGACCTG---TATGCCCCCAAGCCCACTGGCAAGGC 3106
Qy 994 uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeu 1014
Db 3107 CTCCAAGGGAACAACAAAGCAAGGACAGAGACGATCCCAAGCCCGTGAAGCCAGT 3166
Qy 1014 uAspArgAsnAspAsnProProValLeuGlyAsnProGluIleLeuPheAsnAlaTyrVa 1034
Db 3167 GAGAGACAGAGATGAGCGCGGAGCAGAGATGCTCAAGTTC----- 3209
Qy 1034 IThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054
Db 3210 ---AACTGATGAGCAGTATGCCCTGGGAGACAGT---CCCCGATC-----CACCTGCC 3256
Qy 1054 oAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValle 1074
Db 3257 C----- 3257
Qy 1074 uLeuAsnAlaSerThrArgIleuLeuValSerArgAlaLeuAsnAsnArgProle 1094
Db 3258 -CTCACTACCAACCAAGGACGCTGACCTGCGCGCACTATGCTTACTCCCACT 3316
Qy 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111
Db 3317 GCCTTCATTCAGCTGACGACCCCACTCACTCAAGCTCCCAAGACCAAGCCAGTGTGTA 3376
Qy 1111 aGln-CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131
Db 3377 GAGACTGCGACCTGCA-----ACACATTCG----- 3402
Qy 1131 euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeuG 1147
Db 3403 -----TGGGACCGGGGACACCAAGTCCACGGGCTGTGAGCAGT----- 3441
Qy 1147 ILeuPheIleGlnAlaValAlaIleThrLeuAlaThrProProAspHisValValAlP 1167
Db 3442 -----ACTCGAGTACAGCTACCGACCAACCCCA----- 3474
Qy 1167 heAsnVal-GlnArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186
Db 3475 --AAATACCCCAAGCAAGATTAACCTACCGCGCGCTCACTCTC----- 3517
Qy 1187 ValGlnProProGlyProGlyGlyGlyProProPheLeuProSerGlu 1203
Db 3518 ---GGCCACCAAGCCAGGCCCAAGCTGACAGACCC---ATCCCAAGCAG 3562

```

```

ATTORNEY/AGENT INFORMATION:
NAME: No. 5663300and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32660
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 4650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 495..4103
US-08-453-274B-102

Alignment Scores:
Pred. No.: 3,256-46 Length: 4650
Score: 881.50 Matches: 343
Percent Similarity: 38.23% Conservative: 146
Best Local Similarity: 26.82% Mismatches: 452
Query Match: 5.67% Indels: 339
DB: 1 Gaps: 46

US-09-916-849a-3 (1-2923) x US-08-453-274B-102 (1-4650)
Qy 3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeu 22
Db 513 AGCCAGGCGCTGGGGGCAACCGCTACTGCTCCCTCCATGCTGATGCACTGCTC 572
Qy 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42
Db 573 CTGCTGCTCATCCCAAG-----CCAGCCACTGGGTAGTGA 613
Qy 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTyrLeuGlyProSerSerAla 62
Db 614 CAAGTCCGGAAGAAACAGCCCAACCTT---CATTTG-----GAGCTCGC 661
Qy 63 SerAsnLeuTyrLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82
Db 662 AGCCGATATGTTT-----TCCAGATGGGGCACTGTACAGCTAAGGTGGTGC 715
Qy 83 LeuValPro-----HisIleAspGlyLeuArgVal 92
Db 716 CCCGTACTTCGGGTGATGAGCAAGACAGTGCATTTCACCAAGCAAGCTC----- 769
Qy 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
Db 770 -----CATCGACCGTAGGGGCTCCGTAATGCGAAGAACAGCT 808
Qy 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleLeuSerProGluGlyGlyLeuThr 130
Db 809 CCGTGGTATCCCTGCATCTCTGAGATTGAGT----- 841
Qy 131 LeuProGluGluHisProCysLeuValAlaProGluLeuArgCysGlnSerCysValLeu 150
Db 842 -----ATCTATCACAGACTCTGTCAAAATTC----- 868
Qy 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyGlyArg 170
Db 869 ---GAGCCCGCGCTGTAGAGGGCCAGATAGAGT----- 901
Qy 171 ArgIlyArg-AsnValAsnThrAlaProGluInPheGlnProProSerTyrGlnAlaThrVa 190
Db 902 ACAAGACATCAATGACAACA-----CCCAACTGCTCAACAGTACTCTGGGCAT 938
Qy 190 IProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206
Db 959 CCGTGAACACCAACAGTGGGTCACTCTTCCCATCCCGCTGCTTCA----- 1007

```

RESULT 5  
 US-08-453-274B-102  
 ; Sequence 102, Application US/08453274B  
 ; Patent No. 5663300  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 107  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSES: Marshall, O'Toole, Gerstein, Murray & Borum  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/453,274B  
 ; FILING DATE: 30-MAY-1995

QY 206 eAspProbaSpGluGlyGluAlaGlyArgLeuGluTyrThrMetAspAlaLeuPheAspSe 226  
 Db 1008 -GACCGTATCTGCTGCTCCCAAGGCTGTCATCTATGAGTCGACGGTGGCAAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrAlaGluGlu 246  
 Db 1067 GAGAGAGAGAGCA-----CCACAGCTCATTTGATGGCCAA 1102  
 QY 246 uLeuAspArgGluThrThrLeuSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
 Db 1103 CTTGACCGGTATGAGCGCTGGAGCTCTTATGACCTCACTCAAGGTGCAAGATGGCGGAG 1162  
 QY 266 rProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
 Db 1163 CCCCCACGCGCCACAGAGTCCCTGCTGTCGTCACCGCTTGACCAATGACCAAGCC 1222  
 QY 286 pProValPheGluGlnGlnGluTyrTyrGluSerLeuArgGluAsnLeuGluValGlyTyr 306  
 Db 1223 CCCCAGATTGAGCGGCTCTCTATGAGCCGAACATCTGAGAAATAGCCCCCATAGGCCA 1282  
 QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTyr 326  
 Db 1283 CTGCTCATCCAGGTGAAGGCCAATGACTCAACCAAGGTGCCAATGCAAAATCGAATA 1342  
 QY 326 rArgLeuLeuGluGlySerGlySerProSerGluVal-----PheGluIleAs 343  
 Db 1343 CACATTCACACAG-----GGGCCGAAGTTGTAGGCGTCTTCTGACACGGA 1390  
 QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTyr 363  
 Db 1391 CAGGAACACTGGACTTTATCATCTGTTCAGGCCCGGTGACCGTGAAGACCTTAAGCACCT 1450  
 QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCTCACTGCTTGTGTAAGACCAAGGACCAAC-----CCCAAGGTGCCCTGTC 1504  
 QY 383 alaIleValPheLeuSerValGluAspAspAsnAspAsnAlaProGlnPhe----- 399  
 Db 1505 CCAAGTGTGTGACCGTGAAGACATGAATGACATGCCCCCACTGAGATCCGGGG 1564  
 QY 400 -----SerGluYsArgTyrValValGlnValArgGluAspValThrProGly 415  
 Db 1565 CATAGGCTATGACTCATCAAGATGAGATGGCTTAACATCTCAAGATGTGGCAGAGA 1624  
 QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValValHis 435  
 Db 1625 GACAGCTGTGCGCTGTGTCAGGTGTCTGACCGAATGAGGAGAGAAATGACGTGTC-- 1682  
 QY 435 sTyrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGACGATGTGTGCTTCAGCTGCGCAGGCGCAGTGAGACAGCAG 1741  
 QY 455 a-----LeuAspValValSerProLeuAspTyrGluThrThrTyr 468  
 Db 1742 TGACAGAGAGAGAAATTTCTGACAGACTACCAACCCCTGATACATCAAGAGGTCAA 1801  
 QY 468 sGluTyrThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCATTAAGATTGTGGCTGTGACTCTGGCAACCCCTCTCTCCAGCACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerHis 508  
 Db 1862 CTCCTCTC---AAGGTGACGAGGTGTGACGTCAATGACACGCACTGTCTTCACTCAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyrIleuValLeuHisValGly 528  
 Db 1919 TGTCACTGAGAGGTGCTTCCCGGAAAACAACAGCTGTGTGAAGTGAATCTGAGACAC 1978  
 QY 528 naIleAlaAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyrTyrArgLeu----- 544  
 Db 1979 TGCCAGAGATGTGACTCTGCTCTAATGTGAGCTGTGTTACTCTGAGAGCTGAGCC 2038

QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyrPheSerVal 564  
 Db 2039 GCGTGTCAAGAGG-----CTCTTACCATCTCAACCGGAGACTGGAAGATCCAGGT 2089  
 QY 564 lAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyValGluAlaArgAs 584  
 Db 2090 GAAGACATCTCTGATTCGGGACAGCGGAGACCTATGAGTTGAAGGTGTGCGACGCTGA 2149  
 QY 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGCACTCTTACGCTTCAGGGCAACAGCCACTGTCTTGTGAATGTGTCGACATGCAA 2209  
 QY 604 naAspAsnProThrPheThrGlnProGluTyrThrValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TGACATATGACCCAAA----- 2225  
 QY 624 aValGlyThrSerValValThrValSerAlaValaAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 QY 644 rTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGly 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluArgGlnTyrValLeuAl 684  
 Db 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValaAsnValThrAs 704  
 Db 2225 ----- 2225  
 QY 704 palAsnThrHisArgProValPheGlnSerSerHisTyrThrValaAsnValaAsnGluAs 724  
 Db 2226 -----TTATGCTGAGTGTCTCAACTTCTGATGAGAGAA 2263  
 QY 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCCAGCACTGAGTCCAGTGGGCAATGTGATCTGATGATGAGAACAGGGGAGAA 2323  
 QY 744 naIleArgIleThrTyrPheMetGluAspSerIleProGlnPheAspGlyIleAspAlaAspTh 764  
 Db 2324 TGCCAGGTGAGCTCTCACTGAGTGAAGAGCAACGCTGACTTGTATCAAGATGGCAC 2383  
 QY 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSerTyrThrLeuAl 784  
 Db 2384 AGGCACCATCTTATCCAGCTGAGCTTGTATGAGAGCAACAAAGCACTTACCTTCCA 2443  
 QY 784 aIleThrAlaArgAspAsnGlyTyrLeuGlnYsSerAspThrThrTyrLeuGlnIleLe 804  
 Db 2444 GCTGAAGCAGAGATGTGTGGGTGCCCATCGCTCAGCTTACGTTGTGTGATCCATCAA 2503  
 QY 804 uValaAspValaAsnAspAsnAlaProGlnPheLeuAspAspSerTyrGlnGlySerVal 824  
 Db 2504 TGTGCTGACAGAAATGACAAAGCAACCTATATCACTGCCCTTCTC-----AACACCTC 2557  
 QY 824 lTyrGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCACAAAGCTGTGACCCCCCAACACACGTCTTGTGAGACGGTCAAGCAGGTGGACCCA 2617  
 QY 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2618 GGCATTGTGACTGTGGTGTCAATGCGAGCTGATCAACATTCAGTGCAGGTGGCAACCTC-- 2675  
 QY 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValaArgThrLeuAspArgLeuAs 880  
 Db 2676 -TATGACCTCTTCAGATTGGGTCACTTCAAGTGTCCATCACCTTGAAGAGAGAGATTGA 2734  
 QY 880 pArgGluAsnValaAlaGlnTyrValLeuAspArgAlaTyrAlaValaAspGlyGlyMetProPr 900  
 Db 2735 GCGGCGCCACATGGGCTTACACCGCTGTGTGTGAAGTCAATGACCGGCGGCAAGCCCC 2794  
 QY 900 oAlaArgThrProMetGluValThrValThrValLeuAspValaAsnAspAsnProProVa 920



QY 111 ProTpsSerCyArGLeuLeuGlyIleGlyGlyHisLeuSerProGlnGlyIleLeuThr 130  
 Db 809 CCTGTGTATCCCTGCATCCTGATTTGAGGT----- 841  
 QY 131 LeuProGlnGluHisProCySerLeuValaProArgLeuArgCyGlnSerCySerLeu 150  
 Db 842 -----ATCTATCACAGACCTCGGCGAAGATC----- 868  
 QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGlyArgSerProGlnGlySerLeuGlyArg 170  
 Db 869 ---GAGCCCCCGGCTGCTAGAGGCGCAATAGAGT----- 901  
 QY 171 ArgGlyArg-AsnValaSerThrAlaProGlnPheGlnProProSerTyrglnAlaThrVa 190  
 Db 902 ACAAGACATCAATGACACACA---CCCACTTCGCTCAACAGATCACTCTGCGCAT 958  
 QY 190 LProGlnGlnGlnProAlaGlyThr-----ProValAlaSerLeuArgAla 206  
 Db 959 CCTGAGAACCAACATCGGCTCACTTCCCATCCGCTGCGCTCA----- 1007  
 QY 206 eAspProArgGlnGlyGlnAlaGlyArgLeuGlnTyrrThrMetAspAlaLeuPheAspSe 226  
 Db 1008 -GACCGTATGCTGGTCCCAACGATGGGCATCTATGAGCTGAGGTCGAGGCGAGACCA 1066  
 QY 226 rArgSerArgGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlnG 246  
 Db 1067 GGAGGAGAGCA-----CCACAGCTCATGTGATGGGCA 1102  
 QY 246 uLeuAspArgGlnThrTyrrSerThrHisValaPheArgValThrAlaGlnAspHisGlyMe 266  
 Db 1103 CTTGACCTGTAGCCCTGAGACTCCTATGACCTCAACATCAAGGTCGAGATGGCGGAG 1162  
 QY 266 LProArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAspHisAs 286  
 Db 1163 CCCCCACGCGCGAGAGCCCTGCTGGGTGACCGGTGTCACCGTGTGACACCAATGACAAAGC 1222  
 QY 286 PProValPheGlnGlnGlnGlnTyrrTyrrGlySerLeuArgGluAsnLeuGlnValGlyTy 306  
 Db 1223 CCCCAGTTTGAAGCCCTCTATGAGGCCGAACTATCTGAGAAATGACCCCTGAGGCA 1282  
 QY 306 rGlnValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTy 326  
 Db 1283 CTCGGTATCCAGGTGAGGAGGCAATGATCAAGCCAGTGCATGACGAATGATGAT 1342  
 QY 326 rArgLeuLeuGlnGlySerGlySerProSerGlnVal-----PheGlnIleAs 343  
 Db 1343 CACATTCACACAG-----GCCGCCGAAGTTGTGAGGCGTCTTCTTCAGCTGA 1390  
 QY 343 PProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnValGlnSerTy 363  
 Db 1391 CAGGAACACTGACCTTATCACTGTTCAGGCGCCGGTGAACCGTGAAGACCTTAAGCACTT 1450  
 QY 363 rGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCAAGTGTCTGTAGGACCGAGGCAAC-----CCCAAGATGCCCCGTC 1504  
 QY 383 AlaValaPheLeuSerValGlnAspAspAspAspAlaProGlnPhe----- 399  
 Db 1505 CCAGGTGTGTGACCGTGAAGGACATGATGACATGCCCCCACATTCAGATCCGGGG 1564  
 QY 400 -----SerGlnTyrrArgTyrrValValGlnAlaArgGluAspValThrProG 415  
 Db 1565 CATAGGGCTAGTACTCATCAAGATGAGTGGCTTAACATCTCAAGAGATGGCGAGAGA 1624  
 QY 415 ValaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValaHis 435  
 Db 1625 GACAGCTGTGGCCCTGAGCGAGGTGTCTGACCGATATAGAGGAGAGATGACGCTGC-- 1682  
 QY 435 eTyrrSerIleMetSerGlyAsnAlaArgGlyGlnPheTyrrLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGACGATGTGTCCTTCAGCTGCGCAAGCGACTGAGACAGGCGAG 1741  
 QY 455 a-----LeuAspValaSerProLeuAspTyrrGlnThrThr 468

Db 1742 TGACAGCAAGAGAGATATTCCTGACAGACTAACACCCCGCTAGACTACGAAAGTCAA 1801  
 QY 468 eGlnTyrrThrLeuArgValaArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCATGATGATTTGGGCTGTGGACTCTGGCAACCCCACTCTCCAGCACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAspAlaProIlePheValSerTh 508  
 Db 1862 CTCCTC---AAGGTGAGGTGTGGAGCTGAATGACAAACCACTGCTTCACTACAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlyTyrrLeuValLeuHisValG 528  
 Db 1919 TGTCACTGAGGTGCTCTCCGGAACCAACAGCTGTGTAAGATGATTCGATGATCAC 1978  
 QY 528 AlaIleAspAlaAspAlaGlyAspAspAlaArgLeuGlnTyrrArgLeu----- 544  
 Db 1979 TGCAGTATCTGATCTGCTGCTTAATGCTAGCTGTTACTCTCTGAGACCTGAGCC 2038  
 QY 545 -AlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyrrPileserv 564  
 Db 2039 GGCTGCTTAAGGC-----CTTTCACCATCTCAACCGAGCTGAGAGATCAAGT 2089  
 QY 564 LAlaAlaGlnLeuAspArgGlnGlnValaAspPheTyrrSerPheGlyValaGlnAlaArg 584  
 Db 2090 GAAGACATCTCTGATCGGGAACGCGAGAGCTATGATGTAAGGTGTGCGAGCTGA 2149  
 QY 584 PHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValaLeuAspValAs 604  
 Db 2150 CCGGGGACGTCTTAGCTTCCAGGACACAGCCACTGTCTTCAATGTCTGAGCTGCA 2209  
 QY 604 nAspAsnAspProThrPheThrGlnProGlnTyrrThrValaArgLeuAsnAlaAla 624  
 Db 2210 TGACATGACCCCAA----- 2225  
 QY 624 aValGlyThrSerValaThrValaSerAlaValaAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 QY 644 rTyrrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGly 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyrrLeuGlnGlnGlnTyrrValaLeu 684  
 Db 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValaValaSerValThrAs 704  
 Db 2225 ----- 2225  
 QY 704 PAlaAsnThrHisArgProValPheGlnSerSerHisTyrrThrValaAsnValaAsnGluAs 724  
 Db 2226 -----TTATGCTGAGTGGGCTTCAACTTCAAGTATGAGAA 2263  
 QY 724 rArgProAlaGlyThrThrValaValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATCCAGCACTGAGTGCAGTGGGCTAGTGACTGTCAATGATGAGAACAGGGGAGAA 2323  
 QY 744 nAlaArgIleThrTyrrPheMetGlnAspSerIleProGlnPheArgGlyIleAspAlaAspTh 764  
 Db 2324 TGCAGGTGAGCTCTCACTGAGGACAGAACACGCTGACTTGTATCCGAATGGCAC 2383  
 QY 764 rGlyAlaValaThrThrGlnAlaGlnLeuAspTyrrGluAspGlnValaSerTyrrThrLeuAl 784  
 Db 2384 AGGCACATCTCTTCAAGCTGAGCTTGTATCGAGGACCAACAAAGCACTTCACTTCA 2443  
 QY 784 aIleThrAlaArgAspAsnGlyIleProGlnTyrrSerAspThrThrTyrrLeuGlnIle 804  
 Db 2444 GCTGAAGCGAGTGAATGTGTGCTCCACTCTGAGCTTACGTTAGTGTGTACACATCAA 2503  
 QY 804 uValaAspValaAspAspAsnAlaProGlnPheLeuArgAspSerTyrrGlnGlySerVa 824





Db 573 CTCTGGCTTCATCCCGAGG-----CCAGCCACTGGGGTAGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCyProSerSerAla 62  
 Db 614 CAAGGTCCCGGAGGAACGCCCAACACCTT---CATGG-----GAGCTCGC 661  
 QY 63 SerAsnLeuThrLeuThrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82  
 Db 662 AGCCGACTATAGTTT-----TCCAGATGTGGGCACTGTACAACTAGAGGTGGGTGC 715  
 QY 83 LeuValPro-----HisHisValMetGlyLeuVal 92  
 Db 716 CCGGTACTTTCGGGTGATGCAAGACAGGTGACATTTTCCACCAAGGAGCTC----- 769  
 QY 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110  
 Db 770 -----CATGACACCTGAGGGGCTCCGTGATGCAAGAACAGCT 808  
 QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyIleHisLeuSerProGluGlyIleThr 130  
 Db 809 CCTGTGATCCTGTCATCCCTGAGTTTGAAGT----- 841  
 QY 131 LeuProGluGluHisProCysLeuValAspAlaProArgLeuArgCysGlnSerCysIleVal 150  
 Db 842 -----ATCTATCAGACACTCGTCAGAAATGC----- 868  
 QY 151 AlaGluAlaProGlyLeuArgAlaGlyGluArgSerProGluGluSerLeuGlyIleArg 170  
 Db 869 ---GAGCCCGGCTGTATAGGGCCAGATGAAGT----- 901  
 QY 171 ArgIleArg-AspValAsnThrAlaProGluPheGlnProProSerTrpGluAlaThrVal 190  
 Db 902 ACAAGACATCAATGACACACACAC---CCCACTTCGCTCAGCTCAGTCACTCGGCAT 958  
 QY 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206  
 Db 959 CCTGTGAAACCAACATCGGCTCACTTCCCATCCGCTGCTTCA----- 1007  
 QY 206 eAspProAspGluGlyIleValAlaGlyArgLeuGluThrIleMetAspAlaLeuPheAsp 226  
 Db 1008 -GACCGTATGCTGTGCTCCCAAGGTGTGATCTTATGAGCTCAGGTGCAAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 Db 1067 GGAGGAGAACCA-----CCAGCTCATTTGTGAGGCA 1102  
 QY 246 uLeuAspArgGluThrThrSerThrHisValPheArgValThrAlaGluAspHisGly 266  
 Db 1103 CCTGAAACCGGAGCGCTGGAGACTCTTATGACCTCACCACTCAAGGTGCAAGATGGCGCA 1162  
 QY 266 cProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHis 286  
 Db 1163 CCCCCCAGCGGCAACAGATGCTCCTGTGCTGTACACCGTCTTGACACCAATACCAAGC 1222  
 QY 286 pProValPheGluGlnGluGluThrIleValIleValIleValIleValIleValIleVal 306  
 Db 1223 CCCCAGATTTGAGCGGCTCTTATGAGCGGCAACTATCTGAGAAATAGCCCCATAGGCCA 1282  
 QY 306 rGluValIleuThrValAlaArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeu 326  
 Db 1283 CTGGTCACTCAGGTGAGGCAATGACTCAAGACCAAGGTGCCAATGAGAAATACAAAT 1342  
 QY 326 rArgLeuLeuGluGlySerGlyIleSerProSerGluVal-----PheGluIleAs 343  
 Db 1343 CACATTCACACAG-----GGCCCGAAGTTGTAGGCGTCTTCTGACTGGA 1390  
 QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSer 363  
 Db 1391 CAGAAACCTGAGCTTATCACTGTTCAGGGCCGGGTGAGCCGTGAGAGACCTTAAGCACT 1450  
 QY 363 rGlnLeuThrValGluIleAspArgGlnIleArgAspProGlyProArgSerThrThrAl 383

Db 1451 GCGCTTCTCAGTCTTGTATAGACCAAGCAACAC-----CCCAAGATGCCCGTC 1504  
 QY 383 alaValPheLeuSerValGluAspAspAsnAlaProGlnPhe----- 399  
 Db 1505 CAGGTGTGTGACCGTAAAGACATGATATGACATGCCCCCACTTGAATGCCGGG 1564  
 QY 400 -----SerGluValArgGlyThrValGluValArgGluAspValThrProGlu 415  
 Db 1565 CATAGGCTTAGTACTCATCAAGATGGATGGCTAACATCTCAGAGGATGTGCAGAGA 1624  
 QY 415 yAlaProValIleuValValThrAlaSerAspArgAspGlySerAsnAlaValHis 435  
 Db 1625 GACAGCTGTGGCTGTGTGCAGGTGTCTGACCGAGTGTGGAGAGAAATGCACTGTC-- 1682  
 QY 435 rTrpSerIleMetSerGlyAsnAlaArgGlyIleGlnPheThrLeuAspAlaGluThrGly 455  
 Db 1683 -ACCTGTGTGTGGCAGGTATGTGCTTCCAGCTGCGCCAGGCCAGGAGACAGGCA 1741  
 QY 455 a-----LeuAspValIleSerProLeuAspTrpGluThrThr 468  
 Db 1742 TGACAGCAAGAAAGTATTTCTGACACTACACCCCGCTAGACTACAGAAAGTCA 1801  
 QY 468 rGluThrThrLeuValArgAlaGlnAspGlyIleArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCACTGATGATGTGGCTGTGACTCTGGCAACCCCACTCTCCAGCACTTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAlaProIlePheValSerThr 508  
 Db 1862 CTCCTC---AAGGTGCAAGTGTGTGACTCAATACAGCACTGTCTTCACTCAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTrpLeuValLeuHisVal 528  
 Db 1919 TGTCACTGAGTGTGCTTCCCGGAAACCAACAGCTGTGAAGATGTGAGATCAC 1978  
 QY 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluThrArgLeu----- 544  
 Db 1979 TCCAGTATGTGACTCTGTCTTATGTCTAGGTGTGTTACTCTGTGAGCTGAGCC 2038  
 QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTrpIleSerVal 564  
 Db 2039 GGTCTGTAAGGC-----CTTTCACATCTCACCCAGACTGAGAGATCCAGGT 2089  
 QY 564 lAlaAlaGluLeuAspArgGluGluValAspPheTrpSerPheGlyValGluAlaArg 584  
 Db 2090 GAAGCACTCTGTGATCGGGAACAGCGGAGAGCTATGTTGAAGGTGGCGACTGA 2149  
 QY 584 rHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrAlaLeuAspVal 604  
 Db 2150 CCGGAGCTGCTTACCTTCAGGGCAAGCCACTGTCTTTCAAATGTGTGACTGCA 2209  
 QY 604 nAspAsnAspProThrPheThrGlnProGluThrValArgLeuAsnGluAspAla 624  
 Db 2210 TGACAAATGACCCCAA----- 2225  
 QY 624 aValGlyThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleThr 644  
 Db 2225 ----- 2225  
 QY 644 rTrpGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGly 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTrpTrpLeuGluArgGlnTrpValLeuAl 684  
 Db 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValIleAsnValThr 704  
 Db 2225 ----- 2225  
 QY 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTrpThrValAsnValAsnGlu 724  
 Db 2226 -----TTTATGCTAGTGGCTTCAACTTCTCAGTATGAGAA 2263



NAME/KEY: CDS  
LOCATION: 495..4103  
US-08-453-702A-102

Alignment Scores:  
Pred. No.: 3,25e-46 Length: 4650  
Score: 881.50 Matches: 343  
Percent Similarity: 38.23% Conservative: 146  
Best Local Similarity: 26.82% Mismatches: 452  
Query Match: 5.67% Indels: 339  
DB: 2 Gaps: 46

US-09-916-849A-3 (1-2923) x US-08-453-702A-102 (1-4650)

```

QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeuLeu 22
Db 513 AGCCGAGGCGCTGGGGGCAACGGCTACTGCTGCTCCATGCTGCTGCACTGCTGCTC 572
QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42
Db 573 CTGCTGCTCTCATCCCAAG-----CCAGCCCACTGGGGTAGGTGA 613
QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuGlyProSerSerAla 62
Db 614 CAGAGTCCCGAGAGAACGCCAACCCACCTT---CATTTG-----GAGCCTGCG 661
QY 63 SerAsnLeuThrLeuTyThrSerArgCysArgAspAlaGlyThrGlnLeuThrGlyHis 82
Db 662 AGCCGACTATAGTTT-----TCCAGATGTGGGCACTGTACAGTAAAGTGGGTGC 715
QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92
Db 716 CCGGTACTTTCGGGTGATGCAAGACAGGTGACATTTTACACACGAGACCTC----- 769
QY 93 TrpCysProGlnSerGlnAlaHisGlyPro-----LeuProProAlaProGlnGlyCys 110
Db 770 -----CATCAGCCGTAGAGGGCTCCGTGATGCGAAGCAACGACT 808
QY 111 ProTrpSerCysArgLeuLeuGlyGlyGlyHisLeuSerProGlnGlyLeuThr 130
Db 809 CCTGTGTGATCCTGTGATCTCTGAGTTTGAAGT----- 841
QY 131 LeuProGlnGlnHisProCysLeuValProArgLeuArgCysGlnSerCysValLeu 150
Db 842 -----ATCTATCAGACACTCGTGCAGAAATGC----- 868
QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGlyLeuArgSerProGlnGlnSerLeuGlyGlyArg 170
Db 869 ---GAGCCCCGGCTGTAGAGGCCAGATGAAGT----- 901
QY 171 ArgGlyArg-AsnValAsnThrAlaProGlnPheGlnProProSerTyrglnAlaThrVal 190
Db 902 ACAAGCATCATGACACACACA-----CCCACTTGGCTCACAGTACATCTGGCCAT 958
QY 190 LProGlnLeuGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206
Db 959 CCTGTGAAACCAACATCGGCTCACTTCCCATCCCGCTGGCTTCA----- 1007
QY 206 eaSPProAspGlnGlyGlnAlaGlyArgLeuGlnTyThrMetCaspAlaLeuPheAspse 226
Db 1008 -GACCGTATGCTGTGCTCCCAAGGTGGGCACTCTATAGCTCAAGGTGCAGAGAGACCA 1066
QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGln 246
Db 1067 GGAGGAGAGCAA-----CCACAGCTCATTTGTATGGGCAA 1102
QY 246 uLeuAspArgGlnThrIlySerThrHisValPheArgValThrAlaGlnAspHisGlyTyr 266
Db 1103 CCGGACCGGTGACGCTGGGACTCTTAATGACCTCAACATCAGTCAAGGTGCAGAGTGGCGGAG 1162
QY 266 cProArgSerSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286
Db 1163 CCCCCACGGGCAACAGTGCCTGTGCTGCTGTACCGGTGCTTGAACCAATGACAAAGC 1222

```

```

QY 286 pProValPheGlnGlnGlnGlyTyThrGlySerLeuArgGlnAsnLeuGlnValGlyTy 306
Db 1223 CCCCAGTTTACGGGCGCTCTTATAGCCCGGAATCTTGTGAATAGCCCACTAGAGCA 1282
QY 306 rGlnValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTy 326
Db 1283 CTGGGACAGGTGAGGAGCCAAATGACTCAGACCAAGGTGCCAATGAGAAATTCAAAT 1342
QY 326 rArgLeuLeuGlnGlySerGlySerProSerGlnVal-----PheGlnLeuAs 343
Db 1343 CACATTCACACAG-----GCGCCCAAGTTGTAGAGCGCTTCTTCGACTGGA 1390
QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGlnValGlySerTy 363
Db 1391 CAGAAACATGACTTATACCTTACGTTCAGGCGCGGTGACCGTGAACCTTAAGACCTTAC 1450
QY 363 rGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383
Db 1451 GCGCTTCTCAGTGTGCTTAAGGACGAGGACCAAC-----CCCAAGATGCCCGTGC 1504
QY 383 alaValPheLeuSerValGlnAspAspAsnAspAsnAlaProGlnPhe----- 399
Db 1505 CCAAGTGTGTGACCGTGAAGACATGATATACATATGCCCCACCATTAAGATCCGGGG 1564
QY 400 -----SerGlnLeuArgTyThrValGlnValArgGlnAspValThrProG 415
Db 1565 CATAGGCGTATGACTCATCAAGATGAGTGCATACATCTCAGAGAGATGTGCAGAGAA 1624
QY 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValAlaHis 435
Db 1625 GACAGCTGTGGCCCTGTGTGACAGTGTCTGACCGAATGAGGAGAGATGACGCTGC-- 1682
QY 435 eTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyThrLeuAspAlaGlnThrGlyAl 455
Db 1683 -ACTGTGTGTGAGGAGTGAATGTGCTTCCAGCTGCGCCAGCCAGTGAACAGGCGAG 1741
QY 455 a-----LeuAspValAlaSerProLeuAspTyGlnThrThrTy 468
Db 1742 TGACAGCAAGAGAGATATTTCTTCAGACTACACCCGCTAGACTAGAGAGATCAA 1801
QY 468 eGlnTyThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488
Db 1802 AGACTTACACCATTAAGATTTGTGGCTGTGGACTGTGGCAACCCCACTCTCCAGACTAA 1861
QY 488 rGlyLeuValThrValGlnValLeuAspIleAspAsnAlaProIlePheValSerHis 508
Db 1862 CTCCCTC--AAGTGCAGGTGTGACGTCAATGACACGCACTGTCTTCACTCAGAG 1918
QY 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlyTyThrLeuValLeuHisValG 528
Db 1919 TGTCACTAGAGTGGCGCTTCCCGGAACAAACAGCTGTGTGAAGTATGCTGAGATCAC 1978
QY 528 mAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlnTyArgLeu----- 544
Db 1979 TCCAGATGATGTGACTGTGCTTAATGCTGAGCTGTTTACTCTGAGACCTGAGCC 2038
QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTrpIleSerVal 564
Db 2039 GCGTGTCAAGGCG-----CTCTTCACTTCAACCCAGACTGAGAGATCCAGGT 2089
QY 564 lAlaAlaGlnLeuAspArgGlnGlnValAlaAspPheTySerPheGlyValGlnAlaArgAs 584
Db 2090 GAAGACATCTGTGATCGGGAACAGCGGAGAGCTATAGATTGAAGTGTGTGCGAGCTGA 2149
QY 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604
Db 2150 CCGGGGCAAGTCTTACCTTCAAGGGCAAGCACTGTCTTTCATATGTGCTGAGCTGCA 2209
QY 604 nAspAsnAspProThrPheThrGlnProGlnTyThrValArgLeuAsnGlnAspAlaAl 624
Db 2210 TGACAAATACCCCAAA----- 2225

```

OY	624	avalGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIle	644
Db	2225	-----	2225
OY	644	rTyrgInIeThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGly	664
Db	2225	-----	2225
OY	664	yGlyLeuValSerLeuAlaLeuProLeuAspTyrIleGluGluArgGlnTyrValLeuAl	684
Db	2225	-----	2225
OY	684	avalThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValaAsnValThrAs	704
Db	2225	-----	2225
OY	704	pAlaAsnThrAlaArgProValPheGlnSerSerHisTyrThrValaAsnValaAsnGluAs	724
Db	2226	-----TyrArgGTAAGTGGCTAACATTCCTGAGTGAAGTGA	2263
OY	724	PARProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAspThrGlyGluAs	744
Db	2264	CATGCCAGCATCGAATCCAGTGGGAGATGTGACTGTCAATTAATGAAGACAAAGGGGAGAA	2322
OY	744	nAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh	764
Db	2324	TGCCAGGAGCGCTCTCAGTGGACGACAAACGGTGACTTTGTTATCCGAATGGAC	2383
OY	764	rGlyValaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValaSerTyrThrLeuAl	784
Db	2384	AGGCAACCATCTATCCAGCTTGAGCTTGTATGAAGAACAACAGCACTTAACCTTCCA	2443
OY	784	alleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrTyrTyrLeuGluIle	804
Db	2444	GCTGAAGGACAGTGAATGGTGGCTCCACCTGCTCAGCTTACGTATGGTGCATCATCA	2503
OY	804	uValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerVa	824
Db	2504	TGTGTGTGACGAGAAATGACAAAGCAACCTTATCACTGCCCCCTCT-----AACACCTC	2553
OY	824	IlyrGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh	840
Db	2558	TCACAAGCGCTGACACCCCAACACAGCTTGATGTGAACGTCACGCAAGTGGACGCGA	2611
OY	840	rAspArgAspSerGlyLeuAsnGlyArgValaPheTyrThrPheGlnGlyGlyAspAspGl	860
Db	2618	GGACTTGTGACTCTGGTGTCAATGCCGAGCTGATCTACAGCAATTCGACGGTGGCAACCT--	2675
OY	860	yAspGlyAspPheIleValGluSerThrSerGlyIleValaArgThrLeuArgLeuAs	880
Db	2676	-TATGACCTTTCAGATTGGCTACATTCAGGTGCATCACCCCTGGAGAAAGAAATTGA	2733
OY	880	pArgGluAsnValaAlaGlnTyrValaLeuArgAlaTyrAlaValaAspGlyMetProBr	900
Db	2735	GGGGCGGCACCAATGGGCTACACGGCTGTGGTGAAGTCAAGTGAACGGGGCAAGCCCC	2794
OY	900	oAlaArgThrProMetGluValaThrValaThrValaLeuAspValaAsnAspProProVa	920
Db	2795	ACGCTATGGCACAGCTTGCTGCATCTTATATGTCAAATGAGACTTGGCCCAACGGCAGCT	2851
OY	920	lPheGluGln-----AspGluPheAspValaPheValGluGluAsnSerProIleGl	937
Db	2855	GCTGAGAGACCCCTCTGGGCAACAGCCTGAG-----ACCCCGTGA	2891
OY	937	yLeuAlaValaAlaArgValaThrAlaThrAspProAspGluGlyThrAsnAlaGln----	955
Db	2897	TATTGAAT-----GCTGGGATCCAGAAATGAGCCGCTCCAAACAGCGTGG	2944
OY	956	-----IleMetTyrGlnIleValaGluGlyAsnIleProGluValaPheGlnLeuAspIlePh	974
Db	2945	CAACATTCCTCTTGGTGTGGTGGCTAGTGTGTGGTGGCCGTGGCTTGTCTATC-----	2991
OY	974	eSerGlyGluLeuThrAlaLeuVal-----	982

[illegible]



[illegible]

QY	880	paTggluAsnValaIaGlnTyrValleuArgAlaTyrAlaValaAspIleGlyMetProPr	900
Db	2735	GGGGCGCCACATGAGGGCTACCGCCCTGGTGTGTAAGGTCAGTACGCGGGCAAGCCCC	2794
QY	900	caIaArgThrProMetGluValThrValThrValleuAspValaAsnAspAsnProProVa	920
Db	2795	ACGCTATGAGCACAGCCTTGGTCCATCTTATATGTCAATGAGACTCTGGCCAAACCGACGT	2854
QY	920	IphGluGln-----AspGluPheAspValaPheValGluGluAsnSerProIleG	937
Db	2855	GCTGGAAACCTCCTGGGGCCACAGCCTGAGC-----ACGCGCTGGA	2896
QY	937	YleuAlaValaIaArgValaThrAlaThrAspProAspGluGlyThrAsnIaGln-----	955
Db	2897	TATTGACATT-----GCTGGGGATCCGAATATGAGGGCTTCCAGCAGCCTGG	2944
QY	956	---IleMetTyrGlnIleValGluGluAsnIleProGluValaPheGluIleuAspIlePh	974
Db	2945	CAACATTCTCTTGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2996
QY	974	eSerGlyGluLeuThrAlaLeuVal-----	982
Db	2997	-----GCCCTGGCGGTTCTTGTGCGCTACTGACAGACGGAGGCCAAAGTGTTA	3049
QY	983	-----AspLeuAspTyrGluAspArgProGluTyrValle	994
Db	3050	CCAGGCTGTGAAGAGAGACCAAGACCTG---TATGCCCCCAAGCCCACTGGCCAGAGC	3106
QY	994	uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValaIaValaArgleu	1014
Db	3107	CTCCAAAGGAAACAAAGAGCAAGAGCAAGACAGCATCCCAAGGCCCGGTGAAGCCAGT	3166
QY	1014	uAspArgAsnAspAsnProProValleuGlyAsnPheGluIleLeuPheAsnAsnTyrVa	1034
Db	3167	GGAAGACGAGAGTATAGCCCGGCTGTGCAAGTCTCTCAAGTTC-----	3206
QY	1034	IThrAsnArgSerSerPheProGlyGlyAlaIleGlyArgValaProIaIaAspPr	1055
Db	3210	-----AACCTGATGAGCATGCGCCCTGGGGACAGT---CCCGGATC-----CACCTGC	3256
QY	1054	oAspIleSerAspSerLeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValle	1074
Db	3257	C-----	3257
QY	1074	uLeuAsnAlaSerThrGlyGluLeuTyrLeuSerArgAlaLeuAspAsnAsnArgProle	1094
Db	3258	-CTCACTACCCACAGCAGCAGCCTTACCTGTGGCGGCACTATGCTTCACTCCCACT	3318
QY	1094	uGluAlaIleMetSerValleuValSerAspGlyVal-----HisSerValThrAl	1111
Db	3317	GCCTTCATCATCGATGACGCCAGCATCACTCCTCAAGAACCAAGACCAAGGTGTACA	3376
QY	1111	agIn-CyAlaLeuArgValaThrIleIleThrAspGluMetLeuThrHisSerIleThrL	1133
Db	3377	GGAACCTGCACCTGTCA--ACACATTTCG-----	3402
QY	1131	euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeuG	1147
Db	3403	-----TGGGCAACGGGGAGACCAACAGCTCCAGGGCTCTGAGCAGT---	3441
QY	1147	YleuPheIleGlnAlaValaIaAlaThrLeuAlaThrProProAspHisValaValaP	1167
Db	3442	-----ACTCCGACTACAGCTACCGGACCAACCCCCCA-----	3474
QY	1167	heAsnVal-GlnaArgAspThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer	1186
Db	3475	--AATACCCCAAGCAGAGATTACTCACTCCGCGGCTCACTTTC-----	3517
QY	1187	ValGlyGlnProProGlyProGlyGlyGlyProPheLeuProSerGlu	1203
Db	3518	---GGCCACCAAGCCAGGCGCCAGAGCTGTGAGGAGCC---ATCCACAGCACAG	3562

## RESULT 10

PCT-US93-12588-102  
Sequence 102, Application PC/TUS9312588  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: Protocadherin Materials and Methods  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
ADDRESSEE: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12588  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/998,003  
FILING DATE: 29 DEC 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Grete E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4650 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 495..4103  
PCT-US93-12588-102  
Alignment Scores:  
Pred. No.: 3,25e-46 Length: 4650  
Score: 881.50 Matches: 343  
Percent Similarity: 38.23% Conservative: 146  
Best Local Similarity: 26.82% Mismatches: 452  
Query Match: 5.67% Indels: 339  
Gaps: 46  
US-09-916-849a-3 (1-2923) x PCT-US93-12588-102 (1-4650)  
Qy 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22  
Db 513 AGCCAGGCGCTGGGGGCAAGGCTACTGCTGCTCCCTCCATGCTGACCTGCTC 572  
Qy 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42  
Db 573 CTGCTGCTCATCCCAAGG-----CCAGCCCACTGGGTAGTGTGA 613  
Qy 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyThrLeuCysProSerSerAla 62  
Db 614 CAAGTGGCGGAGGAGCAAGCCCAACACCTT---CATTGG-----GAGCCCTGCC 661  
Qy 63 SerLeuLeuThrLeuValThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82  
Db 662 AGCCGACTATGATGTTT-----TCCAGATGTGGGGCAGCTGTACAGTAGAGTGAGTGC 715  
Qy 83 LeuValPro-----HisHisAspGlyLeuValArgVal 92

Db 716 CCCTGACCTTCGGGTGATGCAAGACAGGTGACATTTTCCACCAAGGAGCTTC----- 769  
Qy 93 TrpCysProGluSerGlnValHisIlePro-----LeuProProAlaProGluGlyCys 110  
Db 770 -----CATTCAGCTGTAGAGGGCTCTCGTGAATGCCAGAACCAAGCT 808  
Qy 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGlnGlyIleuThr 130  
Db 809 CCCTGGTGAATCCCTGATCTCGAGGTTTGAGGT----- 841  
Qy 131 LeuProGluGlnHisProCysLeuValAlaProArgLeuArgCysGlnSerCysValLeu 150  
Db 842 -----ATCTATCACAGACTCTGTGCAGAAATGC----- 868  
Qy 151 AlaGlnAlaProGlyLeuArgAlaGlyArgSerProGluGluSerLeuGlyArg 170  
Db 869 ---GAGCCCCCGGCTGTATGAGGGCCAGATGAAAT----- 901  
Qy 171 ArgLeuArg-AsnValAsnThrAlaProGlnPheGlnProProSerTrpGlnAlaThrVal 190  
Db 902 ACAAGACATCAATGACAGACA---CCCAACTTGCTCAACAGTCACTCATCTGGGCAAT 958  
Qy 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206  
Db 959 CCCTGAGAACCAACCAATCGCTCCTCCATCCCGCTGCTTCA----- 1007  
Qy 206 eAspProAspGluGlyGlnAlaGlyArgLeuGlyIleThrMetAspAlaLeuPheAspSe 226  
Db 1008 -GACCGTGAATCTGTGGTCCCAAGGTGGCATCTTATGACTCAGGTGCAGAGGACCA 1066  
Qy 226 ArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
Db 1067 GAGAGGAGAGCA-----CCAGACTCATTTGTATGGGCAA 1102  
Qy 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
Db 1103 CCTGAAACCGTGAAGCGTGGAGCTCTATGACCTCACCATCAAGGTGAGATGGCGGAG 1162  
Qy 266 tProArgSerSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
Db 1163 CCCCCCAGCGGCAACAGATGCTCTGCTGCTGCTGACCGGTCTTGACCAACATGACCAAGC 1222  
Qy 286 pProValPheGluGlnGlnGlnIleuTrpLysGluSerLeuArgGluAsnLeuGlyValGly 306  
Db 1223 CCCCAAGTTTGAAGCGCTCTATGAGGCGGCAATCTATCTGAGATATAGCCCAATGAGCCA 1282  
Qy 306 rGluValLeuLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeu 326  
Db 1283 CTGGTCATCTCAAGTGAAGGCCAATGATCTCAAGCCAAAGTGCACAAATGCAATTA 1342  
Qy 326 rArgLeuLeuGluGlySerGlyLysSerProSerGluVal-----PheGluIleAs 343  
Db 1343 CACATTCACACAG-----GCGCCGAAGTTGTAGAGCGCTCTTCCGACTGGA 1390  
Qy 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTrp 363  
Db 1391 CAGGAACACTGACTATCACTGTTCAGGGGCCGGGTGAGACCGGTGAGAGCTTAAGCAAGCCT 1450  
Qy 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
Db 1451 GCGCTTCTCAGTCTGTCTTGAAGCCAGGACCAAC-----CCCAAGAGTCCCGTGC 1504  
Qy 383 alaValaPheLeuSerValGluAspAspAsnAspAlaProGlnPhe----- 399  
Db 1505 CAGAGTGTGTGACCGTGAAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1564  
Qy 400 -----SerGluValArgGlyValValGlnValAlaArgGluAspValThrProG 415  
Db 1565 CATAGGCTAGTATGATCAATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1624  
Qy 415 yAlaProValLeuArgValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHis 435



Db 1625 GACAGCTGTGGCCCTGGTGCAGTGTCTGACCGAGATGAGGAGAGATGACGCTGTC-- 1682  
 Qy 435 eTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGGAGTGTGATGTGCTTCACGCTGCGCCAGGCGCATGTAGAGAGGACAG 1741  
 Qy 455 a-----LeuAspValIleSerProLeuAspTyArgIleThrIly 468  
 Db 1742 TGACAGAGAGAGAGATATTTCTTCAGACATCAACCCCGCTTACAGAGAGATCA 1801  
 Qy 468 eGluTyThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCAATGAGATGTGGCTGTGACTCTGGCAACCCCGCATCTCCAGCAGCTAA 1861  
 Qy 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508  
 Db 1862 CTCCTTC---AAGGTGAGGTGTGTGAGCTGATGACAGCAGCATCTGCTTCACTCAGAG 1918  
 Qy 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlyTyLeuValLeuHisValG 528  
 Db 1919 TGTCACTGAGGTGCGCTTCCCGAAGAACACAGCGCTGTGAGATGTGCTGAGATCAC 1978  
 Qy 528 nAlaIleAspAlaAspAlaArgIleAspAsnAlaArgLeuGlnTyArgLeu----- 544  
 Db 1979 TGCAGATGATGCTGATCTGCTGCTTAATGCTGAGCTGATTACTCTCTGAGCCTGAGCC 2038  
 Qy 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyPileSerVa 564  
 Db 2039 GGTCTCTAAGGCGC-----CTCTTCAACATCTCACCCGAGACTGAGAGATCTCAGGT 2089  
 Qy 564 lAlaAlaGlyLeuAspArgGlnGluValAspPheTySerPheGlyValGlnAlaArgAs 584  
 Db 2090 GAAGACATCTCTGATGCGGAGAACAGCGGAGAGCTATGATGAAGGTGCGAGCTGTA 2149  
 Qy 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGAGTCTTGAAGCTCCAGGAGCAGCCAGCTGCTTGTCAATGTGCTGAGCTGCA 2209  
 Qy 604 nAspAsnAsnProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TGACATATACCCCAA----- 2225  
 Qy 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 Qy 644 rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225  
 Qy 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyIleLeuGlnArgGlnTyValLeuAl 684  
 Db 2225 ----- 2225  
 Qy 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704  
 Db 2225 ----- 2225  
 Qy 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyThrValAsnValAsnGluAs 724  
 Db 2226 -----TTATGCTGAGTGGCTTACAACTTCTCACTGATGAGAGAA 2263  
 Qy 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCAAGACATGATGCGATGGGATGCTGATGATGATGAGCAAGGGGAGAGAA 2323  
 Qy 744 nAlaArgIleThrTyPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 Db 2324 TCCCAAGGTGAGCTCTCACTGAGAGCAGCAACGGTGACTTTGTTATCCAGATGGCAGC 2383  
 Qy 764 rGlyAlaValThrThrGlnAlaGlyLeuAspTyArgIleAspGlnValSerTyThrLeuAl 784  
 Db 2384 AAGCACCATCTATTCAGCTTGAAGCTTGTATGAGAGCAACAAAGCACTTACCTTCCA 2443

Qy 784 aIleThrAlaArgAspAsnGlyIleProGlnTySerAspThrThrTyLeuGlnIleLe 804  
 Db 2444 GCTGAAGGAGAGATGATGAGGTGCTCCACCTGCTAGCTTACGTTGAGTGCATCATCAA 2503  
 Qy 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuAspAspSerTyArgIleSerVa 824  
 Db 2504 TGTGCTGAGACGAGAAATGACAGACCTTATATCATCTGCCCTTCT-----AACACCTC 2557  
 Qy 824 lTyArgIleAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCACAGCTGTACCCCGACAGACAGCTTGTGTGAGAGGATCAGCAGGAGGACCCGA 2617  
 Qy 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyGlyAspAspG 860  
 Db 2618 GAGCTTGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2675  
 Qy 860 yAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAs 880  
 Db 2676 -PATGATCTTTCAGATTTGGGTCAATTCAGGTGCATCACCTGAGAGAGAGATTA 2734  
 Qy 880 pArgGluAsnValAlaGlnTyThrValLeuArgAlaTyAlaValAspIleGlyMetProP 900  
 Db 2735 GCGGCGCCACATGGCTTACACCGCTGTGTGAGAGGATCAGTACCGGCGAGCCGCC 2794  
 Qy 900 oAlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProProVa 920  
 Db 2795 AGCCTATGGACAGCCTTGTGTGATTTATATGATGATGATGATGATGATGATGATGATG 2854  
 Qy 920 lPheGlnGln-----AspGluPheAspValPheValGlnGluAsnSerProIleG 937  
 Db 2855 GCTGAGAGACCTCTGCGGACAGCAGCTGAGC-----ACGCGGTGGA 2896  
 Qy 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGln----- 955  
 Db 2897 TATTGACATT-----GCTGGGATCCAGAAATATGAGCGCTCCAGAGAGCGTGG 2944  
 Qy 956 -----IleMetTyArgIleValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePh 974  
 Db 2945 CAACATCTCTTGTGGT 2996  
 Qy 974 eSerGlyGlnLeuThrAlaLeuVal----- 982  
 Db 2997 -----GCCCTGCGGCTTGT 3049  
 Qy 983 -----AspLeuAspTyArgIleAspArgProGlnTyValLe 994  
 Db 3050 CAGGCTGTGTAAGAGAGACCAAGACCTG-----TATGCCCAAGCCCAAGTGGCAAGGC 3106  
 Qy 994 uValIleGlnAlaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014  
 Db 3107 CTCCAAGGAGAAACAAAGCAAGAGCAAGAACCAAGTCCCAAGCCCTGAGAGCCAGT 3166  
 Qy 1014 uAspArgAsnAspAsnProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyVa 1034  
 Db 3167 GAGAGACGAGATGAGGCGCGGCTGCAGAAATCCCTCAAGTTC----- 3209  
 Qy 1034 lThrAsnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054  
 Db 3210 -----ACCTGATAGAGATGCGCTTGGGAGACAT-----CCCGGATC-----CACCTGCC 3256  
 Qy 1054 oAspIleSerAspSerLeuThrTySerPheGlnArgGlyAsnGlnLeuSerLeuValLe 1074  
 Db 3257 C----- 3257  
 Qy 1074 uLeuAsnAlaSerThrGlyGlnLeuTyIleUserArgAlaLeuAspAsnAsnArgProLe 1094  
 Db 3258 -CTAACTACCCACAGGAGCGCTGACCTGGCGGCGCATATGCTCTTAATCCCACT 3316  
 Qy 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111  
 Db 3317 GCCTTCATTCAGCTGAGCGCCAGTCAACCTTCAAGCTTCAAGAGAGACAGGAGGTACA 3376

QY 1111 agln-cysalaleuargvalthrilethrhapglumcleturhisertlethr 1131  
 Db 3377 GAACTTCACCTGCA-----ACACATTCG----- 3402  
 QY 1131 euargleugluapmserpro-----gluargpneuuserproleueng 1147  
 Db 3403 -----TGGGCACTGGGGACACACGCTCCAGGCTCTAGAGT----- 3441  
 QY 1147 lyleupheileglnalavalalathrleuathrproproaspheivalvalp 1167  
 Db 3442 -----ACTCCGACTACAGCTACCGACCAACCCCCCA----- 3474  
 QY 1167 healeuval-glunargapthrleupalaarglygluhsileuvalserleuser 1186  
 Db 3475 --AATACCCGAGAGAGAGTACTACCGCCGCGTCACTCTCC----- 3517  
 QY 1187 valglvglnproproglprogllygllyprophleuproserglu 1203  
 Db 3518 ---GCCACACAGCCAGCCAGAGCTGCAAGACC---ATCCAGACAG 3562

RESULT 11  
 PCT-US95-08071-102  
 ; Sequence 102, Application PC/TUS9508071  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Suzuki, Shintaro  
 ; TITLE OF INVENTION: Protocadherin Materials and Methods  
 ; NUMBER OF SEQUENCES: 115  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &  
 ; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/08071  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/12588  
 ; FILING DATE: 23 DEC 1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/998,003  
 ; FILING DATE: 29 DEC 1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta B.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 32149  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 102:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4650 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 495..4103  
 ; PCT-US95-08071-102  
 ;  
 ; Alignment Scores:  
 ; Pred. No.: 3,25e-46 Length: 4650  
 ; Score: 881.50 Matches: 343

Percent Similarity: 38.23% Conservative: 146  
 Best Local Similarity: 26.82% Mismatches: 452  
 Query Match: 5.67% Indels: 339  
 DB: 5 Gaps: 46  
 US-09-916-849a-3 (1-2923) x PCT-US95-08071-102 (1-4650)  
 QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 22  
 Db 513 AGCCCAAGGCGCTGGGGGGAAGCGCTACTGCTGCTCCATGCTGTAGCATGCTGCTC 572  
 QY 23 LeuLeuLeuProProProLeuLeuGlyAepGlnValGlyProCysArgSerLeuGlySer 42  
 Db 573 CTGCTGGCTCCATCCCAAG-----CAAGCCACTCGGTAAGTGA 613  
 QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCysProSerSerAla 62  
 Db 614 CAAGGTCGCGAGAAACAGCCACCAACCTCT---CATTGG-----GAGCTCGC 661  
 QY 63 SerAsnLeuTrpLeuTyThrSerArgCysArgAapAlaGlyThrGluLeuThrGlyHis 82  
 Db 662 AGCCCACTATGTTT-----TCCAGATGTGGGCACTGTACAGCTAGAGTGGTGC 715  
 QY 83 LeuValPro-----HisHisAepGlyLeuArgVal 92  
 Db 716 CCCGTAACCTTCGCGATGATGCAAGACAGTGAATTTTCAACCGAGACCTC----- 769  
 QY 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110  
 Db 770 -----CATGACCGTGAAGGGGCTCCGTGAATGCGAAGACCAAGCT 808  
 QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGlnGlyIysLeuThr 130  
 Db 809 CCTGGTATCCCTGCATCCCTGAGACTTGAAGT----- 841  
 QY 131 LeuProGluGluHisProCysLeuValAProArgLeuArgCysGlnSerCysIysLeu 150  
 Db 842 -----ATCATCACAGACCTCTGTGAGATGC----- 868  
 QY 151 AlaglnAlaProGlyLeuArgAlaGlyIuArgSerProGluGluSerIleuGlyArg 170  
 Db 869 ---GAGCCCGCGGCTGAGAGGGCCAGATGAGT----- 901  
 QY 171 ArgIysArg-AsnValaenthAlaProGlnPheGlnProProSerTyrglnAlaThrVa 190  
 Db 902 ACAAGACATCAATGACACACA---CCCACTGGCTCCACAGATCACTCTGGCAT 958  
 QY 190 lProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaI 206  
 Db 959 CCTGAGAAACCAACATCGGCTCACTCTCCCATCCGCGTGGCTTCA----- 1007  
 QY 206 eAspProAspGluGlyIuAlaGlyArgLeuGluTyThrMetAapAlaLeuPheAsp 226  
 Db 1008 -GACCTGATGCTGTCCCAAGGAGTGGCATCTTATAGCTGAGCGAGGCAAGACCA 1066  
 QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrAlaGluG 246  
 Db 1067 GGAGGAGAAGCA-----CCACAGCTCATTTGATGGGCA 1102  
 QY 246 uLeuAspArgGluThrIysSerThrHisValPheArgValThrAlaGlnAapHisGlyMe 266  
 Db 1103 CCGTGAACCGTAGCGCTGAGACTCTTATAGCTCACTCAAGGTGACAGATGGCGGAG 1162  
 QY 266 tProArgSerAlaLeuAlaThrLeuThrIleLeuValThrAapThrAapHisAs 286  
 Db 1163 CCCCCAAGCGCCAGAGTGCCTGCTCGCTGCTGACCGTCTTACCACTAGACAAAGCC 1222  
 QY 286 pProValPheGluGlnGlnGluTyrgIysGluSerLeuArgIuAsnLeuGluValGlyTy 306  
 Db 1223 CCCCAGTTTGAAGCGGCTCTTATGAGCGCAACTATCTAGAAATAGCCCATAGGCCA 1282  
 QY 306 rGluValLeuThrValArgAlaThrAapGlyAapAlaProProAsnAlaAsnIleLeuTy 326

Db 1283 CTCGTCATCCAGGTCGAGGCAATGACTCAGACCAAGTGCATGCGAATATCATATA 1342  
 Qy 326 rArgLeuGluGluGlySerGlySerProSerGluVal-----PheGluIleAs 343  
 Db 1343 CACATTCACAG-----GCCCGCAAGATTGTAGGCGCTCTTCACAGTGA 1390  
 Qy 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValIleSerTy 363  
 Db 1391 CAGAAACACTGACCTTATCACTGTTCAAGGCCCGGTGACCTGAGACCTTACACACCT 1450  
 Qy 363 rGluLeuThrValGluIleAspArgGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCACATGCTGTCTTAAGACCGAGGCCAAC-----CCCAAGATGCCCGTGC 1504  
 Qy 383 aAlaValPheLeuSerValGluAspAspAspAspAlaProGlnPhe----- 399  
 Db 1505 CCAGGTGTGTGACCGTGAAGACATGACATGACCCCCACCATTTGATGCCGGG 1564  
 Qy 400 -----SerGluValArgTyValValGlnValArgGluAspValThrProG 415  
 Db 1565 CATAGGCTAGTACTCATCAAGATGGATGGCTTACATCAAGATGTGGCAGAGA 1624  
 Qy 415 yAlaProValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValIle 435  
 Db 1625 GACAGCTGTGGCCCTGTGACAGTGTCTGACCGAGATGAGGAGAGATGCAGCTGTC-- 1682  
 Qy 435 sTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyIleuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGTGAGGTGATGTGCTTCACAGTGGCCAGGCCATGTGAGACGGGAG 1741  
 Qy 455 a-----LeuAspValValSerProLeuAspTyArgIleuThrThly 468  
 Db 1742 TGACGCAAGAAAGATATTCCTGACAGATCAACCCCGCTAGCTACGAGAGTCA 1801  
 Qy 468 sGluTyThrIleuArgValArgAlaGlnAspGlyGlyArgProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCATTTGAGTGTGCTGAGCTGCGCAACCCCGCATCTCCAGAGCTCA 1861  
 Qy 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508  
 Db 1862 CTCCTC--AAGGTGAGGTGTGACGTCAATACAAAGCCTGCTTCACTCAGAG 1918  
 Qy 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyIleuValLeuHsValG 528  
 Db 1919 TGTCACTGAGGTGCGCTCCCGGAAACAACAAGCTGTGATGATGTGCGATCAC 1978  
 Qy 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGluTyArgLeu----- 544  
 Db 1979 TGCCAGTATGCTGACTGTGCTCTTAATGTGAGCTGTGTTACTCTCTGAGCCCTGAGCC 2038  
 Qy 545 -AlaGlyValGlnHsAspPheProPheThrIleAsnAsnGlyThrGlyTPIleSerVa 564  
 Db 2039 GCGTCTTAAGGC-----CTCTCAACATCTCAACCCGAGACTGAGATCAGGT 2089  
 Qy 564 lAlaIleGluLeuAspArgGluGluValAspPheTySerPheGlyValGluIleArgAs 584  
 Db 2090 GAAGACATCTCGATCCGGGACAGCGGAGAGCATGATGATGAAGGTGGTGGACGTGA 2149  
 Qy 584 pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGAGGTCTTACCTTCCAGGCAAGCTCACTGCTTGTTCATGTGCTGAGCTGCA 2209  
 Qy 604 nAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TACATGACCCCAAA----- 2225  
 Qy 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHsSerValIleTh 644  
 Db 2225 ----- 2225  
 Qy 644 rTyGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225

Qy 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyIleuLeuGluArgGlnTyValLeuAl 684  
 Db 2225 ----- 2225  
 Qy 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704  
 Db 2225 ----- 2225  
 Qy 704 pAlaAsnThrHsArgProValPheGlnSerSerHsTyTyThrValAsnValAsnGluAs 724  
 Db 2226 -----TTATGCTGAGTGGCTCAACTTCTCAGTATGAGAA 2263  
 Qy 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCCAGCACTGATGCTCAGTGGCAGTGTGATGCTGATGATGAGACCAAGGGGAGAA 2323  
 Qy 744 nAlaArgIleThrTyPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 Db 2324 TGCCAGGTGACCTGCTCAGTGGAGGACCAAGGTGACTTGTATCCAGAAATGGCAC 2383  
 Qy 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyArgIleAspGlnValSerTyThrIleuAl 784  
 Db 2384 AGGCACCATCTTATCCAGCTGAGCTTGTATCCAGAGCAACAAGCACTTACCTTCCA 2443  
 Qy 784 aIleThrAlaArgAspAsnGlyIleProGlnTySerAspTyThrTyIleuGlnIleLe 804  
 Db 2444 GCTGAAGGCAGTGAGATGTGGCTGCCCTCCCTCAGCTTACGTTGCTGCTACATCA 2503  
 Qy 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyGlnIleSerVa 824  
 Db 2504 TGTGCTGAGCAAGAAATGACAAAGCACCTTATATCACTGCCCTTCT--AACACCTC 2557  
 Qy 824 lTyArgIleAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCACAGCTCTGACCCCGCAGACAGCTCTTGTGAGACGCTGACCGAGGTGGAGCGA 2617  
 Qy 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyGlyAspArg 860  
 Db 2618 GACATTGACTGTGTGATGCAATGCCAGCTGATCTACAGATTCAGAGTGGCAACCT-- 2675  
 Qy 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrIleuArgLeuAs 880  
 Db 2676 -TATGACTCTTCAGATGGGTGCACATTCAGGTGCGCATCACTGAGAGAGATGTA 2734  
 Qy 880 pArgGluAsnValAlaGlnTyValLeuArgAlaTyAlaValAspTyGlyMetProP 900  
 Db 2735 GCGGCGCCACATGGGCTTACACGCGCTGTGTGAGAGTCAGTGACCGCGCAAGCCCC 2794  
 Qy 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspProVa 920  
 Db 2795 ACGCTATGGCAAGCCTGTGCTCATCTTTATGTCAATGACATCTGGCCAAACGACGT 2854  
 Qy 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937  
 Db 2855 GCTGAGAACCTCTCTGGGCCACAGCCTTGAC-----ACGCGCTGGA 2896  
 Qy 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955  
 Db 2897 TATGACATTT-----GCTGGGATTCAGAAATGAGCGCTCCAGACGCTGG 2944  
 Qy 956 -----IleMetTyGlnIleValGluGlyAsnIleProGluValPheGlnLeuAspIlePh 974  
 Db 2945 CAACATTCCTTGT 2996  
 Qy 974 eSerGlyGluLeuThrAlaLeuVal----- 982  
 Db 2997 -----GCCCTGGCGGTTCTTGTGCGTACTGACAGACGCGGAGGCCAAAGTGTGA 3049  
 Qy 983 -----AspLeuAspTyArgIleAspArgProGluTyValLe 994  
 Db 3050 CCAGCTGTGAAGAGAGACCAAGACCTG--TATGCCCCCAAGCCAGCTGGCAAGGC 3106

```

QY 994 uValIIeGlnIaThrSerAlaProLeuValSerArgAlaThrValHisValArgLeuLe 1014
DB 3107 CTCGAAGGAGAAACAAAGCAAGGACAGAGCAAGTCCCAAGGCCGCGAAGCCAGT 3166
QY 1014 uAspArgAsnAspAsnProProValLeuGlyAsnPhgGluIleuPheAsnAsnTYrVa 1034
DB 3167 GGAAGACAGAGATGAGCGCGGCTGACAGAGTCCCTCAAGTTC----- 3209
QY 1034 lThrasnArgSerSerSerPheProGlyGlyAlaIleGlyArgValProAlaHisAspPr 1054
DB 3210 ----AACCTGATGAGCGATGCCCTCGGGAGAGT----CCCCGATC-----CACCGGCC 3256
QY 1054 oAspIleSerSerSerLeuThrTYrSerPheGluArgGlyAsnGluLeuSerLeuValLe 1074
DB 3257 C----- 3257
QY 1074 uLeuAsnAlaSerThrGlyGlyLeuLeuLeuSerArgAlaLeuAsnAsnArgProLe 1094
DB 3258 -CTCACTACCCACAGGACAGCCCTGACCTGGCCGCGCACTATGCTTACCTCCACT 3316
QY 1094 uGluAlaIleMetSerValLeuValSerAspGlyVal-----HisSerValThrAl 1111
DB 3317 GCCTTCATCCAGCTGACGCGCCAGTCACTCCTCAGCTCCAGAGAGACGAGGTGTACA 3376
QY 1111 aGln-CysAlaLeuArgValThrIleIleThrAspGluMetLeuThrHisSerIleThrL 1131
DB 3377 GGACCTGCACCTGCAG-----ACACATTCG----- 3402
QY 1131 euArgLeuGluAspMetSerPro-----GluArgPheLeuSerProLeuLeuG 1147
DB 3403 -----TGGGCACCGGGGACACACGCTCCAGGCTCTAGACAGT----- 3441
QY 1147 lYleuPheIIeGlnAlaValAlaIaIaThrLeuAlaThrProProAspHisValValAlP 1167
DB 3442 -----ACTCCGACTACAGTACCGGACGACCAACCCGCCA----- 3474
QY 1167 heAsnVal-GlnAspArgThrAspAlaProGlyGlyHisIleLeuAsnValSerLeuSer 1186
DB 3475 --AAATACCCCAAGACGATTACTACCGCGGCTCACTTCTC----- 3517
QY 1187 ValGlyGlnProGlyProGlyGlyGlyProPheLeuProSerGlu 1203
DB 3518 ---GGCACCAAGCGCGCCAGAGCTGCAGAGACC---ATCCAGACAGCAG 3562

RESULT 12
US-07-998-003A-94
; Sequence 94, Application US/07998003A
; Patent No. 5643781
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 20 South Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/998, 003A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5643781and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 30903

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 495..3572
; US-07-998-003A-94

Alignment Scores:
Pred. No.: 1,81e-45 Length: 4104
Score: 868.50 Matches: 287
Percent Similarity: 40.59% Conservative: 123
Best Local Similarity: 28.42% Mismatches: 362
Query Match: 5.59% Indels: 239
DB: 1 Gaps: 31

US-09-916-849a-3 (1-2923) x US-07-998-003A-94 (1-4104)
QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22
DB 513 AGCCCAAGGCCCTGGGGGCAAGGCTACGCTCCCTCCATGCTGTAAGACTGCTGCTC 572
QY 23 LeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42
DB 573 CTGCTGCTTCATCCCAAG-----CCAGGCCACTGGGTAGTGA 613
QY 43 ArgGlyArgGlySerSerGlyAlaCysAlaPrometGlyThrLeuGlySerSerSerAla 62
DB 614 CAAGTGCAGGAAGAACAGCCACCAACCTT---CATTTG-----GAGCTCGC 661
QY 63 SerAsnLeuThrLeuTYrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82
DB 662 AGCGAATATGTTT-----TCCAGATGTGGGCACTGTACAGTAAAGGTGGTGC 715
QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92
DB 716 CCCGTACTTCGGGTGATGTCAGAGCAGGTGACATTTTCAACCAAGAGACTC----- 769
QY 93 TrpCysProGluSerGlnAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
DB 770 -----CATGACCGTGAAGGGGCTCCGTAATGCCAGAACCAAGCT 808
QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleLeuSerProGlnGlyLeuThr 130
DB 809 CCTGTGATTCCTGTGATCTGTGAGTTTGAAGT----- 841
QY 131 LeuProGluGluHisProCysLeuValaProGlyLeuArgCysGlnSerCysValLeu 150
DB 842 -----ATCTATCACAGACCTCGTGCAGAAATGC----- 868
QY 151 AlaGlnAlaProGlyLeuArgAlaGlyGluArgSerProGluSerLeuGlyGlyArg 170
DB 869 ---GAGCCCCCGGCTGTAAGGGGCGCAAGTAAGAAT----- 901
QY 171 ArgLeuArg-AsnValAsnThrAlaProGlnPheGlnProProSerTYrGlnAlaThrVa 190
DB 902 ACAAGACATCAATGACACACA---CCCAACTTGGCTCACAGGTATCATCTGGCCAT 958
QY 190 lProGluGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaAl 206
DB 959 CCTGGAACACCAACATCGGCTCACTTCCCATCCGCTGCTTCA----- 1007
QY 206 eaAspProAspGluGlyGlyAlaGlyArgLeuGluTYrThrMetArgAlaLeuPheAspse 226
DB 1008 -GACCGTATGCTGTGCCAAGGTGTGAGCTTATGAGCTGCAGGTGCAGAGAGACCA 1066

```

Qy 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 Db 1067 GAGAGAGAACGA-----CCACAGCTCATTTGGATGGGCA 1102  
 Qy 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGlyMe 266  
 Db 1103 CTGTGACCGTGAAGCGCTGGAGCTCTTATGACCTCACAATGAGATGAGCGGCGAG 1162  
 Qy 266 tProArgSerSerAlaLeuAlaThrLeuThrHisLeuValThrAspThrAspHisAs 286  
 Db 1163 CCCCCACGGGCGACGAGTGCCTGTGCTGCTGACCGGTGCTTGAACAATGACACGCGC 1222  
 Qy 286 pProValPheGluGlnGlnGluThrLysGluSerLeuArgLysAsnLeuGluValGlyTy 306  
 Db 1223 CCCCAGATTGAGGGGCCCTCTTATGAGCGGCACTTCTGAGAAATGACCCCATAGGCCA 1282  
 Qy 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnHisLeuTy 326  
 Db 1283 CTGCGTATCCAGGTGAAGGCCAATGACTCAGACCAAGTGCCAATGCAAGAAATCGAATA 1342  
 Qy 326 rArgLeuLeuGluLysSerGlyLysSerProSerGluVal-----PheGluHis 343  
 Db 1343 CACATTCCACAG-----GGCCCGAAGTTGTGAGCGCTTCTTCCAGCTGA 1390  
 Qy 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGluSerTy 363  
 Db 1391 CAGGAACACTGACTTATCATCTGTTCAAGGCGCCGCTGAGCCGTAGAGACCTTAACACCT 1450  
 Qy 363 rGluLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCTCAGTGTCTGTGAAGACCGAGGCCACCA-----CCCAAGAGTGCCTCGTGC 1504  
 Qy 383 aAlaValPheLeuSerValGluAspAspAspAsnAlaProGlnPhe----- 399  
 Db 1505 CCAGGTGTTGTGACCGTGAAGACATGATGAATGACCCGCCACCATGAGATCCGGGG 1564  
 Qy 400 -----SerGluArgGlyValGluValArgGluValAspValThrProG 415  
 Db 1565 CATAGGGCTAGTACTCATGAATGAGATGGATGGCTTAATCATCAGAGATGTGGCAGAGA 1624  
 Qy 415 yAlaProValLeuArgValThrAlaSerAspArgLysGlySerAsnAlaValValHis 435  
 Db 1625 GACACGTGTGGCCCTGTGTGACGTGTGTGACCGAGTGAAGAGAAATGACGTGTC-- 1682  
 Qy 435 sTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyThrLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACCTGTGTGCTGACAGTGAATGTGCTTCCAGCTGCGCAGGCCCACTGAGACAGGCG 1741  
 Qy 455 a-----LeuAspValIleSerProLeuAspTyThrThrTy 468  
 Db 1742 TGACAGCAAGAAAGATTTCTGTGACATACACCCCGCTAAGCTACGAGAAAGTCA 1801  
 Qy 468 rGluTyThrLeuArgValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTTACACCTTGAATGTGTGCTGTGACTGTGCAACCCCGCATCTCCACACCTAA 1861  
 Qy 488 rGlyLeuValThrValGlnValLeuAspHisAsnAspAsnAlaProIlePheValSerTh 508  
 Db 1862 CTCCTC---AAGGTGACAGTGTGTGAGTCAATACACGCACTGTCTTCACTACAGAG 1918  
 Qy 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyThrLeuValLeuHisValG 528  
 Db 1919 TGTCACTAGAGTGCCTTCCCGAAGAAACACACAGCCTGTGATGATGTGCTGATAC 1978  
 Qy 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTyThrGlu----- 544  
 Db 1979 TGCAGTATGTGCTGCTGTAAAGTCAAGCTGATTTACTCTTGAAGACCTGAGCC 2038  
 Qy 545 -AlaGlyValGlyHisAspPheProPheThrHisAsnAsnGlyThrGlyTyThrIleSerVa 564  
 Db 2039 GCGTCTAAGGCG-----CTCTTACACATCTCACCCGAGACTGAGAGATCAGGT 2089

Qy 564 lAlaAlaGluLeuAspArgGluGluValAspPheTySerPheGlyValGluAlaArg 584  
 Db 2090 GAAGACATCTCGATCGGAAACAGCGGAGACTATGATGTGAAGTGTGGCAGCTGA 2149  
 Qy 584 rHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGCAGTCTTGAAGCTTCAGGGGACAGCACTGTCTTGTCAATGTCTGAGTGC 2209  
 Qy 604 nAspAsnAspProThrPheThrGlnProGluTyThrValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TGACATGACCCCA----- 2225  
 Qy 624 aValGlyThrSerValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 Qy 644 rTyThrIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225  
 Qy 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyThrLysLeuGluArgGlyThrValLeuAl 684  
 Db 2225 ----- 2225  
 Qy 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValAlaAsnValThrAs 704  
 Db 2225 ----- 2225  
 Qy 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyThrValAsnValAsnGluAs 724  
 Db 2226 -----TTATGCTGAGTGTCTCAACTTCTCAGTGAAGAGAA 2263  
 Qy 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
 Db 2264 CATGCCAGCATCAGTCCAGTGGCATGTGACTGATCATTTAGTGAAGACAGAGGGAGAA 2323  
 Qy 744 nAlaArgGlyeThrTyPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
 Db 2324 TGCCGAGTGCAGCTCTCAGTGAAGCAGACCAACGCTGACTTGTATTCAGAAATGGCAC 2383  
 Qy 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyThrGluAspGlnValSerTyThrLeuAl 784  
 Db 2384 AGGACACATCTTATCCAGCTGAGCTTATGAGAGCAACAAAGCACTTACCTTCCA 2443  
 Qy 784 aIleThrAlaArgAspAsnGlyTyLeProGlnLysSerAspThrThrTyThrLeuGluIle 804  
 Db 2444 GCTGAAGGAGTGAATGGTGGGTGCCACCTGCTCAGCTTACGTTAGTGTACACATCAA 2503  
 Qy 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSerTyThrGlnLysSerVa 824  
 Db 2504 TGTGCTGACGAGAAATGACAAACCCACTATATCACTGCCCTTCT---AACACCTC 2557  
 Qy 824 lTyThrAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
 Db 2558 TCAACAGCTCTACCCCGACAGCAAGCTTGTGTGAGACGCTACAGCCAGGTGGACCCGA 2617  
 Qy 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyThrPheGlnGlyGlyAspAspG 860  
 Db 2618 GACTTGAATCTGTGTCAATGCCAGCTGATCTACAGATTCAGAGTGGCAACCT-- 2675  
 Qy 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgLysLeuAs 880  
 Db 2676 -TATGACTCTTCAATGTGGTCAATTCAGGTGCCATCACTCGAGAGAAAGATTA 2734  
 Qy 880 pArgGluAsnValAlaGlnTyThrValLeuArgAlaTyAlaValAlaAspLysGlyMetProP 900  
 Db 2735 GCGGCGCACACAGGGGTACACCGCTGTGTGTGAAGGTCAAGAGACCGGCAAGCCCC 2794  
 Qy 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVa 920  
 Db 2795 ACGCTATGACAGACTTGTATCTTATATGCAATGACATCTGTGCCAACGCAAGCT 2854  
 Qy 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleG 937







CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5708143and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 32658  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 94:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4104 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 495..3572  
 US-08-453-695A-94

Alignment Scores:  
 Pred. No.: 1,81e-45 Length: 4104  
 Score: 868.50 Matches: 287  
 Percent Similarity: 40.59% Conserved: 123  
 Best Local Similarity: 28.42% Mismatches: 362  
 Query Match: 5.59% Indels: 239  
 Gaps: 31

US-09-916-849A-3 (1-2923) x US-08-453-695A-94 (1-4104)

```

QY 3 SerProAlaThrGlyValProLeuProThrProProProLeuLeuLeuLeu 22
DB 513 AGCCGACCCCTGGGGGCAACGGCTACTGCTGCTCCATCGCTAGACATGCTCTC 572
QY 23 LeuLeuLeuProProLeuLeuGlyAspGlnValGlyProCysAspSerLeuGlySer 42
DB 573 CTGCTGCTCATCCCAAGC-----CCACGCCACTGGGTAGTGA 613
QY 43 ArgGlyArgGlySerGlyAlaCysAlaProMetGlyTrpLeuCysProSerSerAla 62
DB 614 CAGAGTCCCGAGGAGAACAGCCACCAACCTT---CATTCG-----GAGCTCCG 661
QY 63 SerAsnLeuTrpLeuTyTrpSerArgCysArgAspAlaGlyThrGluLeuThrGlyHis 82
DB 662 AGCGCATATGAGTTT-----TCCAGATGTGGGGCACTGTACCAAGCTAGAGTGGGTGC 715
QY 83 LeuValPro-----HisHisAspGlyLeuArgVal 92
DB 716 CCCGACCTTCGCGCTGGATGGCAAGACGATTTTCAACCCGAGACCTC----- 769
QY 93 TrpCysProGluSerGluAlaHisIlePro-----LeuProProAlaProGluGlyCys 110
DB 770 -----CATCGACCGTGAAGGGCTCCGTGAATGCCAGAACCAAGCT 808
QY 111 ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisIleuSerProGluGlyLeuThr 130
DB 809 CCTGTGGATCCCTGCATCTCTGAGTTTGAGGT----- 841
QY 131 LeuProGluGluHisProCysIleuValAlaProArgLeuArgCysGlnSerCysIleu 150
DB 842 -----ATCTATCAGACACTCTGTGCAAGATGC----- 868
QY 151 AlaGlnAlaProGlyLeuArgAlaGlyIleuArgSerProGluGluSerLeuGlyArg 170
DB 869 ---GAGCCCCCGGCTGTAGAGGGCCAGATGAAGT----- 901
QY 171 ArgIleuArgAsnValAsnThrAlaProGlnPheGlnProProSerTyGlnAlaThrVal 190
DB 902 ACAAGACATCAATGACACACCA---CCCAACTTCCTTCACATCATCTGGCCAT 958
QY 190 LProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

DB 959 CCTGAGAACCAACCAATCGGCTCACTTCCCATCCCGCTGACTCA----- 1007
QY 206 eaAPProAspGluGlyGluAlaGlyArgLeuGlyTrpThrMetAspAlaLeuPheAsp 226
DB 1008 -GACCGTGAATGCTGGTCCCAACGGTGTGGCATCTTATGACTGCAGGTGGCAAGACCA 1066
QY 226 rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrAlaGlu 246
DB 1067 GGAGAGAGAACCA-----CCACACTCATTTGTGATGGCAA 1102
QY 246 uLeuAspArgGluThrLysSerThrHisValPheArgValThrAlaGlnAspHisGly 266
DB 1103 CCGTGAACCGTGAACCGTGGAGCTCTATATGACTCACCATTCAAGATGATGCGGAG 1162
QY 266 tProArgSerSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHis 286
DB 1163 CCCCCAGGCGCAACAGATGCTGCTGGGTGCACCGGTCTTGACCCATACACACAGC 1222
QY 286 pProValPheGluGlnGluGlyTrpIleGluSerLeuArgIleuAsnLeuGluValGly 306
DB 1223 CCCCAGTTTGAAGCGGCGCTCTATAGAGCGCAACTATCTGAGAAATAGCCCCATAGGCCA 1282
QY 306 rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeu 326
DB 1283 CTGGTCACTCAAGGTGAAGGCCAATGACTCAGACCAAGGTGCCAATGACAAATCAATA 1342
QY 326 rArgLeuLeuGluGlySerGlyLysSerProSerGluVal-----PheGluIleAs 343
DB 1343 CACATTCACACAG-----GGCCCGAAGTTGTGAAGCGCTTCTTCGACTGGA 1390
QY 343 pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGluValGlySer 363
DB 1391 CAGGAACACTGACCTATCATCTGTCAGGGCCCGGTGGACCGTGAAGACTTAAGACACCT 1450
QY 363 rGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSerThrAla 383
DB 1451 GCGCTTCTCAGTCTTGTCTAAGACCCAGGACCAAC-----CCCAAGATGCCCGTGC 1504
QY 383 AlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPhe----- 399
DB 1505 CCAAGTGTGTGACCGTGAAGACATGAAATGCAATGCCCCCAACATTGAGATCCGGGG 1564
QY 400 -----SerGluIleArgTyValValGlnValArgIleuAspValThrProG 415
DB 1565 CATAGGGCTAGTGACTCATCAAGATGGAGTGGTAACTCAAGAGATGTGGCAGAGGA 1624
QY 415 yAlaProValLeuAspValThrAlaSerAspArgAspArgGlySerAsnAlaValHis 435
DB 1625 GACAGCTGTGGCCCTGTGCAGGTCTTGAACGAGTAGAGGAGAGAAATGCAAGCTGTC-- 1682
QY 435 sTySerIleMetSerGlyAsnAlaArgGlyGlnPheTyTrpLeuAspAlaGlnThrGlyAl 455
DB 1683 -ACCTGTGTGTGGAGGTGATGTGCTTCCAGCTGCCCGCCAGGCGCAAGTGAAGACGGCAG 1741
QY 455 a-----LeuAspValValSerProLeuAspTyTrpGluThrTrp 468
DB 1742 TGACAGCAAGAAAGATATTTCTGTGACACTACACCCCGCTAGACTACAGAAAGTCAA 1801
QY 468 sGluTyTrpThrLeuAspValArgAlaGlnAspGlyGlyArgProProLeuSerAsnValSe 488
DB 1802 AGACTACACCATTTGATTTGGCTGTGACCTTGGCAACCCCACTCTCCAGACTTA 1861
QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508
DB 1862 CTCCTCC---AAGTGTGAGGTGTGACCTCATATGACACGACCTGTCTTCACTCAGAG 1918
QY 508 rProPheGlnAlaThrValLeuGluSerValProLeuGlyTyTrpLeuValLeuHisVal 528
DB 1919 TGTCACTGAGTGTGCGCTTCCCGAAACAAACAAGCGCTGTGAAGTGAATTGTGAGATCAC 1978
QY 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTrpArgLeu----- 544
DB 1979 TGCAGTGAATGCTGACTGTGGCTTAATGTGAGCTGTGTTACTCTCTGAGACCTGAGCC 2038

```

```

QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTyrIleSerVa 564
Db 2039 GGCTGCTAAGGCGC-----CTCTTCAACATCTTCAACCCGAGACTGAGAGATCCAGCT 2089
QY 564 1AlaAlaGlyLeuAspArgGlyGluValAlaAspPheTyrSerPheGlyValGluAlaArgAs 584
Db 2090 GAACAACATCTCTGGATCGGGAACAGCGGAGAGCATATGATTGAAGTGGTGGCAGCTGA 2149
QY 584 PHiGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValIleAspValAs 604
Db 2150 CCGGGGCGAGCTGCTGAGCTCCAGGGGACAGCATCTGCTTCAATGCTGAGCTGCA 2209
QY 604 nAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsnGluAspAlaAl 624
Db 2210 TGACATGACCCCA----- 2225
QY 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644
Db 2225 ----- 2225
QY 644 rTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyGly 664
Db 2225 ----- 2225
QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTyrIleGluGluArgGlnTyrValIleuAl 684
Db 2225 ----- 2225
QY 684 aValThrAlaSerAspGlyThrAspGlnAspThrAlaGlnIleValIleAsnValThrAs 704
Db 2225 ----- 2225
QY 704 pAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAs 724
Db 2226 -----TTTATGCTGAGTGGCTACCACTTCTCAAGATGAGACA 2263
QY 724 pArgProAlaGlyThrThrValValIleuIleSerAlaThrAspGluAspThrGlyGluAs 744
Db 2264 CATGCGACACTGAGTCCAGTGGCATGTGATCTGTCAATGATGAGACAAAGGGGAGACA 2323
QY 744 nAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764
Db 2324 TGCCCAAGGTGACGCTCTCGATGAGACAGAACAGCGTACTTTTATTCAGAAATGGCAC 2383
QY 764 rGlyValAlaValThrThrGlnAlaGluLeuAspTyrGlnAspGlnValSerTyrThrLeuAl 784
Db 2384 AGGCACCATCTATCTCAAGCTTGAAGCTTGTATGAGAGAACAAAGACCTTACACCTTCCA 2443
QY 784 aIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyrLeuGluIleLe 804
Db 2444 GCTGAAGGACAGATGATGGTGGCTCCACCTGCTCAGCTTACGTTGATGATCACCATCA 2503
QY 804 uValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerVa 824
Db 2504 TGTGCTGACGAGATGACAAACAGCACTATATCACTGCCCTCT-----AACCACTC 2557
QY 824 lTyrGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840
Db 2558 TCACAAGCTGCTGACCCCAACACAGCTTGTGTGATGAGACGCTGACGCTGAGCCGA 2617
QY 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyIleAspAspGl 860
Db 2618 GGAATTGTGACTCTGGGTGTCATGCGAGCTGATCTACAGCATTGACGAGTGGCAACCTC 2675
QY 860 yAspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAs 880
Db 2676 -TATGACTCTTCCAGATGGGTGACATTCAGTGTGCACCTGAGAGAGAGATTGA 2734
QY 880 pArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAlaAspIleGlyMetProPr 900
Db 2735 GGGGGGCAACAGGAGCTACACGCGTGTGTGTAAGTCACTGACCGGGGCAAGCCGCC 2794

```

```

QY 900 oAlaArgThrProMetGluValThrValThrValIleAspValAlaAspAsnProProVa 920
Db 2795 ACCCTATGGACAGCCTTGCTCATCTTATGTCATATGATGAGACTCTGGCCAAAGCGAGCT 2854
QY 920 lPheGluGln-----AspGluPheAspValPheValGluGluAsnSerProIleGl 937
Db 2855 GCTGAGAGACCTCTGAGCCACAGCAGCCTGAGC-----ACGCCGCTGGA 2896
QY 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGln----- 955
Db 2897 TATTGACATT-----CTGGGATTCAGATATAGCCGCTCCAGACGCTGG 2944
QY 956 -----lIleMetTyrGlnIleValGluGly 963
Db 2945 CAACATCTCTTGTGTGTGTGCTGCTGCT 2972

```

# RESULT 15

US-08-268-161A-94

Sequence 94, Application US/08268161A

Patent No. 5798224

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 115

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 233 South Wacker, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/268,161A

FILING DATE: June 27, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Young J. Suh

REGISTRATION NUMBER: P-41,337

REFERENCE/DOCKET NUMBER: 27866/32149

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 94:

SEQUENCE CHARACTERISTICS:

LENGTH: 4104 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 495..3572

US-08-268-161A-94

Alignment Scores:

Pred. No.: 1,81e-45 Length: 4104

Score: 868.50 Matches: 287

Percent Similarity: 40.59% Conservative: 123

Best Local Similarity: 28.42% Mismatches: 362

Query Match: 5.59% Indels: 239

DB: 1 Gaps: 31

US-09-916-849A-3 (1-2923) x US-08-268-161A-94 (1-4104)

QY 3 SerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeuLeuLeu 22

Db 513 AGCCGAGCCCTGGGGGCAACGGCTACTGCTCCCTCCATGCTGTAGCACTGCTGCTC 572  
 QY LeuLeuLeuProProLeuLeuGlyAspGlnValGlyProCysArgSerLeuGlySer 42  
 Db 573 CTGCTGCTCCATCCCAAG-----CCAGGCCACTGGGTAGTGA 613  
 QY ArgGlyArgGlySerSerGlyAlaCysAlaProMetGlyTrpLeuCysProSerSerAla 62  
 Db 614 CAAGATGCCGAGAGAACAGCCACCCACCTC---CATTGG-----GAGCCCTGCG 661  
 QY SerSerLeuTrpLeuTrpThrSerArgCysArgSerAlaGlyThrGluLeuThrGlyHis 82  
 Db 662 AGCCGACTATGAGTTT-----TCCAGATGTGGGGCACTGTACAGCTAGAGGTGGGTGC 715  
 QY LeuValPro-----HisHisAspGlyLeuArgVal 92  
 Db 716 CCGGTACTTCCGGGTGATGAGCAAGAGGTGACATTTTACACCAAGACCTC----- 769  
 QY TrpCysProGluSerGlnAlaHisGlyPro-----LeuProProAlaProGluGlyCys 110  
 Db 770 -----CATCGACCGTGAAGGGGCTCCGTGAATGCCAGAACAGCT 808  
 QY ProTrpSerCysArgLeuLeuGlyIleGlyGlyHisLeuSerProGlnGlyIleLeuThr 130  
 Db 809 CCTGTGATCCTCGCATCCTGAGATTGAGT----- 841  
 QY LeuProGluGlnHisProCysLeuValAlaProArgLeuArgCysGlnSerCysIleLeu 150  
 Db 842 -----ATCTATCACAGACCTGTGTCAGAAATCC----- 868  
 QY AlaGlnAlaProGlyLeuArgAlaGlyGlyIleArgSerProGluGlnSerLeuGlyArg 170  
 Db 869 ---GACCCCGCGCTGAGAGGGCCAGATAGAGT----- 901  
 QY ArgGlyArg-AsnValAsnThrAlaProGlnPheGlnProProSerTrpGlnAlaThrVal 190  
 Db 902 ACAAGACATCAATGACACACCA---CCCACTTGGCTCACAGTCATCACTTGGCCAT 958  
 QY ProGluAsnGlnProAlaGlyThr-----ProValAlaSerLeuArgAlaIle 206  
 Db 959 CCTGTGAACACCAACATCGGCTCACCTTCCCATCCCGCTGCTTCA----- 1007  
 QY eAspProAspGlnGlyGlyAlaGlyArgLeuGlyTrpMetAspAlaLeuPheAspSe 226  
 Db 1008 -GACCTGTATGCTGTGCCCAACGGTGTGCATCTTATGAGCTGACAGTGGCAGAGACCA 1066  
 QY rArgSerAsnGlnPhePheSerLeuAspProValThrGlyAlaValThrThrAlaGlu 246  
 Db 1067 GGAGGAGAACCA-----CCACAGCTCATTTGTATGGGCA 1102  
 QY uLeuAspArgGluThrIleSerThrHisValPheArgValThrAlaGlnAspHisGlyMet 266  
 Db 1103 CCGGACCGGTGACGGCTGAGCATCTCTTCCCATCCCGCTGCTTCA----- 1162  
 QY cProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAspThrAsnAspHisAs 286  
 Db 1163 CCCCCCAGCGGCAACAGTCCCTGCTGCGTGTACCGTCTTGACCAACATGACCAACCC 1222  
 QY pProValPheGlnGlnGlnGlyTrpIleSerLeuAspArgGluAsnLeuGlnValGlyTrp 306  
 Db 1223 CCCCAGATTGAGCGGCTCTTATGAGCCGCAACTATGTGAGAAATGACCCCAATGAGCCA 1282  
 QY rGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsnAlaAsnIleLeuTrp 326  
 Db 1283 CTCGTGCATCCAGGTGAGGCCAATGACTCAGACCAAGTGCAGAAATGCAAAAT 1342  
 QY rArgLeuLeuGlnGlySerGlySerProSerGluVal-----PheGluIleAs 343  
 Db 1343 GACATTCACACAG-----GCGCCGAAAGTGTGAGCGCTTCTTCGACTGGA 1390  
 QY pProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlnValGluSerTrp 363  
 Db 1391 GAGGAACACTGGACTTATCACTGTTCAGGGCCCGGTGACCGTGAAGACCTTAGCACCTCT 1450

QY rGluLeuThrValGlnAlaSerAspGlnGlyArgAspProGlyProArgSerThrThrAl 383  
 Db 1451 GCGCTTCAAGCTCTTGTCTAAGAGCCAGGACCAAC-----CCCAAGATGCCGTGC 1504  
 QY AlaValPheLeuSerValGluAspAspAsnAsnAlaProGlnPhe----- 399  
 Db 1505 CCAGGTGTGTGACCGGTGAAGACATGATATGACATGCCCCCACCATTGAGATCCGGGG 1564  
 QY 400 -----SerGluAspArgTrpValGlnValAlaArgGluAspValThrProG 415  
 Db 1565 CATAGGCTAGTGACTCATCAAGATGAGATGCTTACATCTCAGAGAGATGTGCAAGAGA 1624  
 QY 415 ValProValLeuArgValThrAlaSerAspArgAspArgGlySerAsnAlaValHis 435  
 Db 1625 GACAGCTGTGCCCTGCTGTACAGGTGTCTGACCGAGATGAGGAGAGATGACGCTG-- 1682  
 QY sTrpSerIleMetSerGlyAsnAlaArgGlyGlnPheTrpLeuAspAlaGlnThrGlyAl 455  
 Db 1683 -ACGTGTGTGTGAGGAGGTGATGTCCTTCCAGCTGGGCCAGGCCAGTGAAGACAGGCA 1741  
 QY 455 a-----LeuAspValAlaSerProLeuAspTrpGluThrThr 468  
 Db 1742 TGACAGCAAGAAAGTATTTCTGACAGACTACCCCGCTAGACTACAGAGAGGTCAA 1801  
 QY sGluTrpThrLeuArgValArgAlaGlnAspGlyArgProProLeuSerAsnValSe 488  
 Db 1802 AGACTACACCATTTGAGATTTGGCTGTGACTCTGCAACCCCGCATCTTCCAGCATTA 1861  
 QY 488 rGlyLeuValThrValGlnValLeuAspIleAsnAspAsnAlaProIlePheValSerTh 508  
 Db 1862 CTCCTCTC---AAGGTGAGAGTGTGTGACCTCATATGACACGACCTGTCTTCACTCAGAG 1918  
 QY 508 rProPheGlnAlaThrValLeuGlnSerValProLeuGlyTrpLeuValLeuHisValG 528  
 Db 1919 TGTCACTGAGTGCCTTCCCGGAAACAAAGCCGTGTGAAGTGTGAGATGTGACATCAC 1978  
 QY 528 nAlaIleAspAlaAspAlaGlyAspAsnAlaArgLeuGlyTrpArgLeu----- 544  
 Db 1979 TCCAGTATGTGACTGCTGCTGCTTAATCTGAGCTGTGTTCTCTGTGAGCTGAGCC 2038  
 QY 545 -AlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGlyTrpIleSerVal 564  
 Db 2039 GCGTCTTAAGGCG-----CTCTTCAACATCTCACCCGAGACTGGAGATTCACAGGT 2089  
 QY 564 lAlaIleGluLeuAspArgGluGlnValAspPheTrpSerPheGlyValGlnAlaArgSe 584  
 Db 2090 GAAGACATCTCTGATCGGAAACAGCGGAGAGCTATGATGAAAGTGTGGCAGCTGA 2149  
 QY pHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrValLeuAspValAs 604  
 Db 2150 CCGGGGAGCTCTTACCTTCCAGGCAACGCACTGTCTTGTCAATGTCTGACTGCA 2209  
 QY 604 nAspAsnAspProThrPheThrGlnProGluTrpValArgLeuAsnGluAspAlaAl 624  
 Db 2210 TGACATATGCCCAAA----- 2225  
 QY 624 aValGlyThrSerValValThrValSerAlaValAspArgAspAlaHisSerValIleTh 644  
 Db 2225 ----- 2225  
 QY 644 rTyroGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSerGlnSerGlyG 664  
 Db 2225 ----- 2225  
 QY 664 yGlyLeuValSerLeuAlaLeuProLeuAspTrpTrpIleGlnGlnGlyTrpValLeuAl 684  
 Db 2225 ----- 2225  
 QY 684 aValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValValAsnValThrAs 704  
 Db 2225 ----- 2225

QY 704 pAlaSerThrHisArgProValPheGlnSerSerHisTyrThrValAsnValAsnGluAs 724  
Db 2226 -----TTTATGCTGAGTGGCTACACACTTCTCAGTATGAGAA 2263  
QY 724 pArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAspThrGlyGluAs 744  
Db 2264 CATGCCAGCTGAGTCCAGTGGAGGAGAGAGACAGCGTGAAGCTTGTATCCAGAAATGGCAC 2323  
QY 744 nAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIleAspAlaAspTh 764  
Db 2324 TGCCCAAGTGAAGTCTCAGTGGAGAGAGACAGCGTGAAGCTTGTATCCAGAAATGGCAC 2383  
QY 764 rGlyAlaValThrThrGlnAlaGluLeuAspTyrGlnAspGlnValSerTyrThrLeuAl 784  
Db 2384 AGGCACCATCTTACCGCTGAGCTTGAATCGAGACCAACAAAGCACCTTACCTTCCA 2443  
QY 784 aIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyrLeuGlnIleLe 804  
Db 2444 GCTGAAGCGATGGATGGTGGCGTCCACTCGCTCAGCTTACGTTGGTGTCCACCATCAA 2503  
QY 804 uValAsnAspValAlaAsnAspAsnAlaProGlnPheLeuArgAspSerTyrGlnGlySerVa 824  
Db 2504 TGTGCTGAGAGAGATGACAAAGCACCTTATATCATCGCCCTTCT-----AACACTC 2557  
QY 824 lTyrGluAspValProProPheThr-----SerValLeuGlnIleSerAlaTh 840  
Db 2558 TCACAACTGCTGACCCCGACACGCTTGTGAGACGTCAGCCAGTGGCGACCGCA 2617  
QY 840 rAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGl 860  
Db 2618 GGACTTGTACTGTGTCTCAATGCCAGCTGATCTACAGCATGTCAGGTGGCAACCT-- 2675  
QY 860 yAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAs 880  
Db 2676 -TATGACTCTTCCAGATTTGGGTCAATTCAGTGCATCATCCCTGAGAAAGAGATTGA 2734  
QY 880 pArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetProPr 900  
Db 2735 GCGGCGCACCATGGGCTTACCGCTGTGTGTGAGAGTCAGTACCGCGCAAGCCCC 2794  
QY 900 oAlaArgThrProMetGluValThrValThrValLeuAspValAlaAsnAspAsnProProVa 920  
Db 2795 ACGCTATGGCACAGCTTGGTTCATCTTTATGTCAATGAGACTGTGGCCAAACGACGCT 2854  
QY 920 lPheGlnGln-----AspGluPheAspValPheValGlnGluAsnSerProIleGl 937  
Db 2855 GCTGAGAGACCTCTCGGGCCACAGCTGAC-----ACGCGCTGGA 2896  
QY 937 yLeuAlaValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGln----- 955  
Db 2897 TATTGACATT-----GCTGGGAGATCCAGAAATATGAGCGCTCAAGCAGCGTGG 2944  
QY 956 ----lMetTyrGlnIleValGlnGly 963  
Db 2945 CAACATTCTCTTGTGTGTGTGGT 2972

Search completed: February 14, 2004, 05:45:23  
Job time : 597 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 14, 2004, 03:42:22, Search time 1948 Seconds  
(Without alignments)  
5527.340 Million cell updates/sec

Title: US-09-916-849a-3  
Perfect score: 15545  
Sequence: 1 MMSPATGVLPPLPPPLLL.....AGTVDESSGSEPLFFNFLH 2923

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-Q/cgn2\_1/USPTO.spool/US09916849.rnat.11022004.143818.21031/arp.query.fasta\_1.3079  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=BIOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USERS=US09916849 @CGN 1.1 1417 @rnat.11022004.143818.21031  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	15545	100.0	8772	US-09-788-711A-3
2	15545	100.0	10531	US-10-225-567A-523
3	15518.5	99.8	8871	US-09-788-711A-1
4	12339	79.4	24370	US-10-292-798-931
5	8781.5	56.5	9045	US-09-737-149-1
6	8781.5	56.5	11389	US-09-728-711-4
7	8781.5	56.5	11389	US-10-295-027-749
8	8781.5	56.5	11389	US-10-240-145-20
9	8781.5	56.5	11389	US-10-241-220-51
10	8781.5	56.5	11389	US-10-225-567A-443
11	8736.5	56.2	9087	US-10-150-811-13
12	8736.5	56.2	9087	US-10-131-409-13
13	7836.5	50.4	11965	US-10-123-409-13
14	7766	50.0	11648	US-10-149-819-42
15	3958	25.5	2332	US-09-764-870-569
16	3958	25.5	2332	US-10-125-540-569
17	3818	24.6	2291	US-10-176-847-99
18	3787	24.4	2603	US-09-843-856-1
19	3734	24.0	4152	US-09-925-300-359
20	3701	23.8	2695	US-10-264-237-636
21	3526	22.7	2077	US-09-764-870-570
22	3526	22.7	2077	US-10-125-540-570
23	3318.5	21.3	3492	US-10-150-811-67
24	3318.5	21.3	3492	US-10-131-409-67
25	3317	21.3	3493	US-10-150-811-68
26	3317	21.3	3493	US-10-131-409-68
27	2070.5	13.3	11002	US-10-017-161-1095
28	1732.5	11.1	15603	US-10-114-153-63
29	1520.5	9.8	14756	US-10-295-027-1106
30	1520.5	9.8	14756	US-10-160-758-5
31	1520.5	9.8	14756	US-10-060-036-134
32	1460	9.4	1376	US-10-029-386-22961
33	1451	9.3	925	US-09-764-870-70
34	1451	9.3	925	US-10-125-540-70
35	1443	9.3	14334	US-10-084-817-227
36	1432	9.2	14749	US-10-210-120-72
37	1423.5	9.2	9294	US-10-174-677-112
38	1415.5	9.1	13700	US-10-085-198-21
39	1338.5	8.6	14336	US-09-970-944-3
40	1338.5	8.6	14336	US-10-160-758-6
41	1338.5	8.6	14336	US-10-160-758-7
42	1331	8.4	14330	US-10-062-674-1800
43	1306.5	8.4	11073	US-10-085-117-113
44	1306.5	8.4	11073	US-10-160-758-2
45	1305	8.4	10065	US-10-085-117-110

## ALIGNMENTS

Sequence 3, Appl1  
Sequence 523, App  
Sequence 1, Appl1  
Sequence 931, App  
Sequence 1, Appl1  
Sequence 4, Appl1  
Sequence 749, App  
Sequence 20, App  
Sequence 443, App  
Sequence 13, Appl1  
Sequence 655, App  
Sequence 42, Appl1  
Sequence 569, App  
Sequence 569, App  
Sequence 99, Appl1  
Sequence 359, App  
Sequence 636, App  
Sequence 570, App  
Sequence 67, Appl1  
Sequence 67, Appl1  
Sequence 68, Appl1  
Sequence 1095, App  
Sequence 63, Appl1  
Sequence 1106, App  
Sequence 5, Appl1  
Sequence 134, App  
Sequence 2261, A  
Sequence 70, Appl1  
Sequence 227, App  
Sequence 112, App  
Sequence 21, Appl1  
Sequence 3, Appl1  
Sequence 6, Appl1  
Sequence 7, Appl1  
Sequence 1800, App  
Sequence 113, App  
Sequence 2, Appl1  
Sequence 110, App

RESULT 1  
US-09-788-711A-3  
Sequence 3, Application US/09788711A  
Patent No. US20020058328A1  
GENERAL INFORMATION:  
APPLICANT: Tania Tamsin Testa  
TITLE OF INVENTION: NOVEL COMPOUNDS  
FILE REFERENCE: GP-30225  
CURRENT APPLICATION NUMBER: US/09/788, 711A  
CURRENT FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 0004196.2  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 8772  
TYPE: DNA  
ORGANISM: HOMO SAPIENS  
US-09-788-711A-3  
Alignment Scores:  
Pred. No.: 0  
Score: 15545.00  
Percent Similarity: 100.00%  
Length: 8772  
Matches: 2923  
Conservative: 0

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

QY	1	LeIaASerProAlaThhGlyValProleuProthProProProleuLeuLeu	20
Db	1	ATGGGAAACCCGGACCAACGGAGCTCCCTCCCAACGCGCGCGCTCTGTGTG	60
QY	21	LeuLeuLeuLeuLeuProProProleuLeuGlyAapGlnValGlyProCyArGSerLeu	40
Db	61	TTGCTGTGTGTCTCCGCCGCCCACTATTGGGAGACCAAGTGGGGCCCTGTGTTCTTG	120
QY	41	GlySerArgGlyYArgGlySerSerGlyValaCysAlaPrometGlyTTPLeuCyEProSer	60
Db	121	GGGTCCAGAGGAGCGAGGCTCTTGGGGGGCTGTGGCCCCCAATGGGTGTGTCTGTCAACC	180
QY	61	SerAlaSerAsmLeuTTPLeuTyThrSerArgCysArgAspAlaGlyThhGluLeuThr	80
Db	181	TCAGGCTGAAACTCTGTGCTTACACACAGCGGCTCAGGAGATGGGGGCACTGAGTGACT	240
QY	81	GlyHisLeuValProHisHisAspGlyLeuArgValTTPCyEProGluSerGluAlaHis	100
Db	241	GGCCACCTGGTACCCCAACACGATGCGCTGAGGGTTTGATGTCCGAATCCGAGGCCAT	300
QY	101	ILeProleuProProAlaProGluGlyCysProTTPSerCysArgLeuLeuGlyYLeGly	120
Db	301	ATTCCCTTACCACCAAGCTCTGTGAAGGTGGCCCTGAGCTGTGGCTCTGGGCAATTGA	360
QY	121	GlyHisLeuSerProGlnGlyYLeuLeuThrLeuProGluGluHisProCysLeuLeuVala	140
Db	361	GGCCACCTTTCCCAACAGGGCAAGTCAACACTGCCGAGAGACACCTGTGCTTAAAGCT	420
QY	141	ProArgLeuArgCysGlnSerCysLeuLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu	160
Db	421	CCAGGCTCAAGTGCAGTCCGTGCAAGGCTGGCAAGGCCGCCCGGCTCAGGGGAGGGAA	480
QY	161	ArgSerProGluGluSerLeuGlyGlyArgArgGlyArgArgenValAsnThrAlaProGln	180
Db	481	AGGTACCAAGAAAGATCTCCCTGGGTGGGTGCGAAAGAAATGTAATACAGCCCCCAG	540
QY	181	PheGlnProProSerTyrGlnAlaThrValProGluAenGlnProAlaGlyThhProVal	200
Db	541	TTCCAGCCCCCAGCTAACAGGCGCACAGTGGCGGAGAACAGCCAGCGAGCACCCCTGTT	600
QY	201	AlaSerLeuArgAlaIleAspProAspGluGlyGluAlaGlyValGluGluGluTyrThhMet	220
Db	601	GCATCCCTGAGGGCCATGACCCCGAGCAGGGGTGAGGAGGTGCACCTGAGTCAACCAATG	660
QY	221	AspAlaLeuPheAspSerArgSerArgGlnPhePheSerLeuAspProValThhGlyAla	240
Db	661	GATGCCCTCTTGAAGACCGCGCTCCAACCAAGTCTCTCCCTGGAGCCCACTGACTGTGTCA	720
QY	241	ValThrThrAlaGluGluLeuAspArgGluThrIlySerThrHisValPheArgValThr	260
Db	721	GTAACACACGCCGAGAGAGCTGTGATGCTGAGACCAAGAGACCCCAAGTCTTCAGGGTCAAG	780
QY	261	AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThhLeuThrIleLeuValThr	280
Db	781	GGCGAGAGACAGCGCATGCCGCCGAGAGATGCCCTGTGCTACCTACCATCTTGGTAACT	840
QY	281	AspThrAsnAspHisAspProValPheGluGlnGlnIleTyrIlyGluSerLeuArgGlu	300
Db	841	GACACCAACAGACATGACCTGTGTTCAGACGACGAGATMACAGAGAGCTCAGGGAG	900
QY	301	AsnLeuGlnValGlyTyrGluValLeuThrValArgAlaThhAspGlyAspAlaProPro	320
Db	901	AACCTGGAGCTTGGCTATAGAGTGTCACTGTCAAGGCCACCGAAGGTGATGTCCTCC	960
QY	321	AsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyIlySerProSerGluValPhe	340
Db	961	AATGCCAAATATCTGTACCGCTGTGAGAGGGGTCTGGGGGAGGCCCTCTGAAATCTTT	1020

QY	341	GIUUIEAPPROAAGSERGIYVAIIIEATGTHRAAGLYPROVALAPAPAGGIUINVAI	360
DB	1021	GAGATACACCTCGCTCGGGGATATCCAAACCGTGAGCCCTGTGGATCCGGAGAGAGTG	1080
QY	361	GIUSERYTGINLEUTHRAIGUALSAPSPGNGIYARGAAPPROGIYPROARGSER	380
DB	1081	GAATCTTACACGACTGACGGATAGAGGCAAGTACACGAGGOTCGGAGCCCGGGCTCGAGT	1140
QY	381	THTRHRAIAAIVAIPIHEUSEVAIGIUBAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP	400
DB	1141	ACCAACGCGCGTGTTCCTTCCTGCTGAGAGATACATGATATATATATATATATATAT	1200
QY	401	GIUUYBARGTYRVAIVAGINVAIARGIUBAPVALTHPRGIIYALAPROVALIENARG	420
DB	1201	GAGAGACCCCTATGTGTGCTCAGGTGAGGAGATGTGACTCCAGGGGCCCCAGTACTCGA	1260
QY	421	VAITHRALASERAPAPAGAPLYEGIYSERANALAVAIVALHISYRSERIIEMESER	440
DB	1261	GTCAACGCTCGATTCACAGACAGAGGGAGCATCCCTGGTGGCACTATACATCATAGT	1320
QY	441	GIYANALIAARGIYGINPHERYLEUBAPALAGINTHRGIYALAEUBAPVALISER	460
DB	1321	GGCAATGCTCGGGAGACAGTTTATCTGATAGCCACGCTGAGCTCTGAGATGTGTAGC	1380
QY	461	PROLEUAPRYTGINUTHRTHRYLSEGLUTYRTHLEUARGVALARGALAGINAPGIYGI	480
DB	1381	CCTCTTGACTATGAGACGACCAAGAGTACACCTTACGGGGGAGACAGAGATGTGTGC	1440
QY	481	ARGPROPLEUSERSENVALSERGIYLEUVALTHYRVALGINVALLEUAPRIEANAP	500
DB	1441	CGTCCCCCACTCTTAAATGTCTCGGGTGGTGACATACAGGCTCGGATATCAAGAC	1500
QY	501	ASNALAPROIIPEPHEVALSERTHPROPHAGINLAETHYRVALEUGIUSERVALPROLEU	520
DB	1501	AATGCCCCCACTTCTGTACAGACCCCTTCCAGGCTACTGTCTCGAGAGGGTCCCTTA	1560
QY	521	GIYTRYLEUVALLEUHIISVALGINALAIIEAPALAPAPALAGIYAPAPANALARGLEU	540
DB	1561	GGCTAACCTGGTCTCCATGCTCCAGGCTATGCAACCTATGTGTGCAAGGCCGCGCTG	1620
QY	541	GIUYTRYARGLEUALAGIYVALGIYHISAPSPHEPROBETHRIIEANAPNGIYTHGCIY	560
DB	1621	GAATACCGCTTGTCTGGGGTGGGACATGACTTCCCTTCCACTACATCAATAGGACAGGC	1680
QY	561	TRIPIESERVALAIAAGIULEUBAPARGIUGIULVALAPPHYRSERPHEGIYVAI	580
DB	1681	TGGATCTCTGTGGCTGTGMACTGACCGGAGGAGATGATTCTTCAAGCTTGGGGTA	1740
QY	581	GIUUIAARGAPPIIAGIYTHPRCALAEUTHRALASERVALSERVALTHYVAL	600
DB	1741	GAAGTCCAGACCAATGGACTCCAGCACTCACTCCGCGCAATGTGCAGGGTACGTGC	1800
QY	601	LEUBAPVALAENAPAPAPAPROPTHETHRGINPROGIUYRTHYRTHYVALARGLEUBN	620
DB	1801	CTGATGTGCACGACACATATCCACTTTRACCAACCAAGATACACAGTCCGCTCAT	1860
QY	621	GIUBAPALAAIYALGIYTHSERVALIYTHYRVALISERVALAVALAPAPARGPALAHIS	640
DB	1861	GAGGATGACGCTGTGGGACCAAGGTGGTGAAGGTGTCAAGCTGTGGACCGGTGATCAT	1920
QY	641	SERVALIIEHTTRYRGIIIEHTHSEGIYASNTHRARGANAPPIESERIIEHTHSE	660
DB	1921	AGTGTACATCACTTACMAATCAACAGTGGCAATCTGANAACCGCTTCTCATACCAAC	1980
QY	661	GINSERGIYGIYLEUVALSERLEUALAENUPROLEUAPRYTYLSEUGIUAIGIN	680
DB	1981	CAAAATGTGTGGGCTGTATCCCTTGCCCTGCACTGACATCAAACTTGAACGGGACG	2040
QY	681	TYRYVALLEUALAVALTHRALASERAPGIYTHYRGINAPPTRIIAGINIIIEVALI	700
DB	2041	TATGTGTGGCTGTATCCGCTCCGANTGGCACTGGCAGGACAGGCAAGATTTGTGTG	2100

QY 701 AsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 Db 2101 AAGTCAACCGACCGCCAAACCCATCGCTCTTTCAGAGCTCCACATATACAGTGAAT 2160  
 QY 722 ValAsnGlnAspArgProAlaGlyThrThrValValLeuLysSerAlaThrAspGluAsp 740  
 Db 2161 GTTAAAGAGACCGGCGGAGGACCAACGGTGTGTATGACGGCAACGGATGAGAC 2220  
 QY 741 ThrGlyGlnAsnAlaArgLysThrTyrPheMetGlnAspSerLysProGlnPheArgLys 760  
 Db 2221 ACAGGTGAGATGCCCCGATCACCTTATCATGAGGACAGCATCCCGCATGCCCATC 2280  
 QY 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGlnLeuAspTyrGluAspGlnValSer 780  
 Db 2281 GATGACGACACCGGGGCTGTCAACCCAGCTGAGGTGATCAACAAGCAAGGTCT 2340  
 QY 781 TyrThrLeuAlaLeuThrAlaArgAspAsnGlyLeuProGlnLysSerAspThrThrTyr 800  
 Db 2341 TACACCTGGCCATTAAGCTCTCGGACATGGCAATCCCAAGAGTCCACACACCTAC 2400  
 QY 801 LeuGlnLysLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db 2401 CTGGAGATCTCGGTGAACGACGTAATGACATGCCCTCAGTTCCTGGGAGCTCTAC 2460  
 QY 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnLysSerAlaThr 840  
 Db 2461 CAGGGACGTGTATAGATGATGATGCCACCTTCACTAGCGTCTCGACGATCTCAGCACCT 2520  
 QY 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspArgGly 860  
 Db 2521 GATCGTGAATCTGAGCTTAATGACAGGGCTTCTTACACCTCCAGAGGCGACGATGGA 2580  
 QY 861 AspGlyAspPheLysValGlnSerThrSerGlyLysValArgThrLeuArgArgLeuAsp 880  
 Db 2581 GACGGGACTTTATTTGTGAGTCAAGTCAAGCATGTGTGAGACGTTACGAGAGCTGAT 2640  
 QY 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspLysGlyMetProPro 900  
 Db 2641 CAGAGAAAGGTGGCCAGATATGTCTTGGCGGCAATAGCAGTGAAGGAGGATGCCCA 2700  
 QY 901 AlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProProVal 920  
 Db 2701 GCCCGACACTTATGAGATGACATGCTGTGTGATGTGATATACATATCCCTGATC 2760  
 QY 921 PheGlnGlnAspGluPheAspValPheValGlnGluAsnSerProLysLeuAlaVal 940  
 Db 2761 TTTGACACAGATGAGTTGATGTGTTGTGAGAGAGAACAGCCCATTTGGGCTAGCCG 2820  
 QY 941 AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnLysMetTyrGlnLys 960  
 Db 2821 GCCCGGCTCAACGCCCTGACCCCGATGAAGGACCAATGCCAGATTTATGTACAGATT 2880  
 QY 961 ValGlnGlyAsnLysProGlnValPheGlnLeuAspLysPheSerGlyGlnLeuThrAla 980  
 Db 2881 GTGGAGGGCAATCTCTGAGGTCTTCCAGCTGACATCTTCTCCGGAGAGCTGACAGCC 2940  
 QY 981 LeuValAspLeuAspTyrGlnAspArgProGluTyrValLeuValLysGlnAlaThrSer 1000  
 Db 2941 CTGGTGAAGCTTGAATACAGAGACCGGCGCTGAGTATGCTGATCCAGGCAAGCTGA 3000  
 QY 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsnPro 1020  
 Db 3001 GCCCTCTGTGTGCGGGCTGACATGTCACGTCCGCTCTTGAACCGCATATGACACCCA 3060  
 QY 1021 ProValLeuGlyAsnPheGlnLysLeuPheAsnAsnTyrValThrAsnArgSerSerSer 1040  
 Db 3061 CCAAGTCTGGGCAACTTTGATCTCTTTTCAACAATATGTCAACAAATGCTCAGACAC 3120  
 QY 1041 PheProGlyGlyValAlaLysArgValProAlaHisAspProAspLysSerAspSerLeu 1060  
 Db 3121 TTCCTGGGGGTGCAATTGGCCGAGTACCTGCCCATGACCTGTATCTCAATATGATCTG 3180  
 QY 1061 ThrTyrSerPheGlnArgLysGlnLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080

Db 3181 ACTTACAGCTTTGAGCGGGGAATGAACTCACCTGCTCTCTCAATCTCCACGGGT 3240  
 QY 1081 GlnLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGlnAlaLysSerVal 1100  
 Db 3241 GAGCTGAAGCTAAGCGCGCACTGACACAACCGGCTCTGTAGAGCCATCATAGCGGTG 3300  
 QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysValAlaLeuArgValThrLysLe 1120  
 Db 3301 CTGGTGTACAGACGGGTACACAGGCTGACCGGCTGAGTGGCGCTGTGTGACATCATC 3360  
 QY 1121 ThrAspGluMetLeuThrHisSerLysThrLeuArgLeuGluAspMetSerProGluArg 1140  
 Db 3361 ACCGATGAGATCTCAACCAAGCATACCGCTGGCCCTGAGAGAAATGTCAACCGGCGC 3420  
 QY 1141 PheLeuSerProLeuLeuGlyLeuPheLysGlnAlaValAlaAlaThrLeuAlaThrPro 1160  
 Db 3421 TTCTGTACACATCTATAGGCTCTTCAATCCAGCGGTGGCCGCAAGTGGCCACGGCA 3480  
 QY 1161 ProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisLysLe 1180  
 Db 3481 CCGACACAGTGTGTGTCTTCAACGTACAGCGGACACCGACGCCCGGGGGCCACATC 3540  
 QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyTyrGlyGlyProProPheLeu 1200  
 Db 3541 CTCAACGTGAGCTGTCCGGTGGCCAGCGCCAGGGCGCGGGCGCGCCCTTCTC 3600  
 QY 1201 ProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuThrAlaLysSer 1220  
 Db 3601 CCTCTAGAGACTCAGAGAGCGCTTATACCTCAACCGAGCTCTGACAGCCATCTCG 3660  
 QY 1221 AlaGlnArgValLeuProPheAspAsnLysCysLeuArgGluProCysGlnAsnTyr 1240  
 Db 3661 GCACAGCGGTGTCTCCCTTGAACGACAAACATCTGCTGGGAGACCTCTGGAGAACTAC 3720  
 QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheLysLeuSerSer 1260  
 Db 3721 ATGCCCTGTGTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780  
 QY 1261 ValLeuPheArgProLysIleProValGlyGlyLeuAsnArgCysArgCysProProGlyPhe 1280  
 Db 3781 GTGCTCTTCCGGCCATCAACCCCTCGAGAGGCTGCGCTGCGCTGCGCCCGGCTTC 3840  
 QY 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 Db 3841 ACGGTGACTACTGTGAGAACGAGGTGACCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 3900  
 QY 1301 GlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 Db 3901 GGGCGCTCCCGAGCGCGAGGGCGGCTACACCTGCTGTGTGTGTGTGTGTGTGTGTGT 3960  
 QY 1321 GlnHisCysGlnValSerLysArgSerGlyArgCysThrProGlyValCysLysAsnGly 1340  
 Db 3961 GAGCACTGTGAGTATAGTCTGCTCAGGCGCTGTGACCCCGGAGTGTCTGCAAGATGG 4020  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAspPhe 1360  
 Db 4021 GGCACTGTGTCAACTCTGT 4080  
 QY 1361 GlnLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheLysThrPhe 1380  
 Db 4081 GAGAGCCCTACTGTGCAAGTACCAACGCGCAGCTTCCCGGCACTCTTCAATCACTTT 4140  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGlnArg 1400  
 Db 4141 CCGGCGCTCGCGCAGCTTTCATCTTCACTTCACTTGGCTCTGTGTGTGTGTGTGTGTGT 4200  
 QY 1401 AspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlnLysHisAspPheValAlaLeuGln 1420  
 Db 4201 GACGGGT 4260  
 QY 1421 ValLysGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlnSerThrThrThrValSer 1440



Db 4261 GTGATCCAGAGCAGATCCAGCTTCTCTGCGAGGAGTACACCAACGAGTGGC 4320  
Qy 1441 ProbetaValProglyValSerAspGlyValIleThrPheIleThrValGlnIleuValTyrTyr 1460  
Db 4321 CCATTCGCGCCGAGAGAGTCAAGTATGCGCAGTGGCACTACGAGTGCATTAATCTAC 4380  
Qy 1461 AsnIleuProleuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnIleValAla 1480  
Db 4381 AATAAGCAGCTGTTGGGTCAAGACAGGGCTCCCAAGAGGCCATCAGAGCAGAAAGTGGCT 4440  
Qy 1481 ValValIleThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
Db 4441 GTGGTGAACCTGGTATGGCTGTGACACAGAGTGGCTTGGCTTGGATCTGCTGGG 4500  
Qy 1501 AsnIleuSerCysAlaAlaGlnGlyIleThrGlnGlySerIleValSerLeuAspLeuThr 1520  
Db 4501 AACTACTCTGCTGCTGCGCAGAGGACCCAGAGGTGGCAGAAAGATCTGTGATCTGACG 4560  
Qy 1521 GlyProleuLeuLeuGlyValProAspLeuProGlySerPheProValArgMetArg 1540  
Db 4561 GGGGCCCTGCTACTAGGGGGGGTGCCTGACCTGGCCGAGAGCTTCCAGTCCGATGGG 4620  
Qy 1541 GlnPheValGlyCysMetArgPheLeuGlnValAspSerArgIleIleAspMetAlaAsp 1560  
Db 4621 CAGTTCGGTGGCTGATCGGAACTGCAAGTGGACAGCCGGCACTAGACATGGCTGAC 4680  
Qy 1561 PheIleAlaAspAsnGlyIleThrValProGlyCysProAlaValIleValAspValCysAspSer 1580  
Db 4681 TTCAATTCGCAACATAGGACCGTGGCTGCTGGCTGCGCAAGAAAGAGTGTGACAGC 4740  
Qy 1581 AsnThrCysValAsnGlyValIleThrCysValAsnGlnIleAspAlaPheSerCysGlyCys 1600  
Db 4741 AACACTTCCCAATAGGGGGGCACTTGGCTGGTGAACAGTGGAGCGCTTACAGTGGAGTGC 4800  
Qy 1601 ProleuGlyPheGlyValIleSerCysAlaGlnGlnMetAlaAspProGlnIlePheLeu 1620  
Db 4801 CCCCTGGGCTTTGGGGGAGAGAGTGGCGCCAGAAATGGCCAAATCCACAGACTTCCG 4860  
Qy 1621 GlySerSerLeuValAlaIleThrIleGlyLeuSerLeuProIleSerGlyProThrTyrLeu 1640  
Db 4861 GGCAGACGCTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 4920  
Qy 1641 SerLeuMetPheArgThrArgGlnIleAspArgValIleLeuLeuGlnAlaIleThrArgGly 1660  
Db 4921 AGCCTCATGTTCCGACCGGCGCAGGCGCAGCGGTGCTGCTGCAAGGCATCCACGAGGG 4980  
Qy 1661 ArgSerThrIleThrLeuGlnIleuArgGlnGlnIleValMetLeuSerValGlnGlyIleThr 1680  
Db 4981 CGCAGCACCATCACTTACAGCTACAGAGGAGCGCAGTATGCTGAGCGTGGAGGGCA 5040  
Qy 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAspArgValAspThrIle 1700  
Db 5041 GGGGCTTCAAGGCT 5100  
Qy 1701 HisAlaGlnIleuAlaLeuGlyValAspArgGlyIleProGlyIleHisAlaIleLeuSerPheAsp 1720  
Db 5101 CATGCAAGCTGGCACTGGGAGCAGCGGGGGGCTGGGCGCAGCATCTGCTCTTCCAT 5160  
Qy 1721 TyrGlyGlnGlnIleArgAlaGlnIleValMetLeuGlyProArgLeuHisGlyLeuHis 1740  
Db 5161 TATGGGAGCAGAGAGCAGAGGAGCAGTGGGCCCCGGGCTGATGCTGACCTGGAC 5220  
Qy 1741 AsnIleThrValGlyValIleProGlyProAlaGlyValAlaAlaArgGlyPheArgGly 1760  
Db 5221 AACATACAGTGGGGGGAATACCTGGGCGAGCGGGGGTGGCCCGTGGCTTGGGGG 5280  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGlyGlnValAsnSerLeuAspProSer 1780  
Db 5281 TGTGTTCAGAGGTGTGGGTGAGCATACGCAAGGGGGTTAAACAGCTGTGATCCAGC 5340  
Qy 1781 HisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5341 CATGGGAGAGATCAACGTGAGAGAGGCTTACGCTGACCTGTGATCTCAAC 5400

Qy 1801 ProCysProAlaAsnSerIleCysSerAspAspThrAspSerIleSerCysSerCysAsp 1820  
Db 5401 CCGTGTCTCTTAACAGCTATTTACAGACAGCATGGGAGAGCTATTTCTGACCTGTGAT 5460  
Qy 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGlnIleGln 1840  
Db 5461 CAGAGTTACTATGTGTACAACTGATTAATGTGTGTGACCTGAACCCGTTGAGCAGCAG 5520  
Qy 1841 SerValCysThrArgIleProSerAlaProHisGlyIleThrCysGlyCysProProAsn 1860  
Db 5521 TGTGTGTGATCCGCAAGCGCAAGTGGCCCCCAAGGCTTATACCTGCAAGTGTCCCCAAT 5580  
Qy 1861 TyrLeuGlyProTyrCysGlnIleThrArgIleAspGlnProCysProArgGlyIleThrGly 1880  
Db 5581 TACCTGGGCAATCTGTGAGACCAAGATTGACCAAGCTTGTCCCTGGCTGGGGGA 5640  
Qy 1881 HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn 1900  
Db 5641 CATCCCAATGTGGCCCATGCACTGATGTGACAGAAAGGCTTGAACCAAGCTGCAAC 5700  
Qy 1901 LysThrSerGlyGlyCysHisCysIleGlnIleThrArgProProGlyIleSerProThr 1920  
Db 5701 AAGACAGCGGAGAGTCCATGCAAGAGAACCACTACCGGCCCCAGGCAAGCCCAAC 5760  
Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgValCysAspProGlyAsp 1940  
Db 5761 TGCTCTGTGTGATCTGATCCCAAGGCTCTTGTTCAGAGTCTGTGACCTGAGAGAT 5820  
Qy 1941 GlyGlnCysProCysIleProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
Db 5821 GGCAGAGTCAATGCAAGCAGAGTGTGATCGGGCGTCAAGTGTGACCGCTGTGACACCT 5880  
Qy 1961 PheAlaGlnValIleThrIleAsnGlyCysGlnValAsnIleThrAspSerCysProArgAlaIle 1980  
Db 5881 TTGTGTAGGTCAACCAATAGGTGTGAAGTGAATTAATGACACTGCCACAGAGCATT 5940  
Qy 1981 GlnAlaGlyIleThrIleProArgThrArgPheGlyLeuProAlaAlaIleProCysPro 2000  
Db 5941 GAGGCTGGAGTGGTGGCCCGGTACCGCTTGGGGCTGCTGCTGCTGCTGCTGCTGCTG 6000  
Qy 2001 LysGlySerPheGlyIleThrAlaValArgHisCysAspArgIleHisArgGlyIleThrLeuPro 2020  
Db 6001 AAGGCTCTTTTGGAGCTGTGTGGCAGCTGATGATGACAGAGGGGTGCTCCCCCA 6060  
Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuValGlyPheAlaGlnArgLeu 2040  
Db 6061 AACTCTTCACTGACAGTCAATCACTTCTCAAGACTGAAGGCTTGGCTGAGCGGCTA 6120  
Qy 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnIleuAlaLeuLeuLeuArg 2060  
Db 6121 CAGCGAATAGTCAAGCTTAAGTCAAGGCGCTCCAGCAGCTAAGCCCTGCTCTCGC 6180  
Qy 2061 AsnAlaThrGlnIleThrAlaGlyTyrPheGlySerAspValIleValAlaIleThrGlnLeu 2080  
Db 6181 AACGCCAGCAGCAGCAGAGCTGTGCTTGGGAGCAGCAGCTGAGGTGGCTTACAGCTG 6240  
Qy 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
Db 6241 GCAAGGGGCTGTGGCCCAAGAGACCCAGGGGGCTTTGGGCTGTGTGCCACAG 6300  
Qy 2101 AspValHisPheThrGlnAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6301 GACGTGCACTTCACTGAAGATCTGCTCGGGTGGGAGCGGCTCTGAGCAGAGCCAAAC 6360  
Qy 2121 LysArgHisThrGlyIleuValIleGlnIleThrGlnGlyValIleThrAlaIleThrLeuGlnHis 2140  
Db 6361 AAGGGGCACTGGAGCTGATTCAGACAGCAGAGGGTGGACCGGCTGTGCTCTCAGAGC 6420  
Qy 2141 TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisIleThrTyrLeuSerProPhe 2160  
Db 6421 TATGAGGCTATCGCAATGGCTTGGCCCAAGACATGGGGCACTTACTTAAGCCCTTC 6480

QY 2161 ThrTlValThrProLeuValIleSerValValArgLeuAspLysGlyAsnPheAla 2180  
DB 6481 ACCATGGTCAAGCCCAACATTTGTCATCTCGTAGTGGCTTGGACAAGGAACCTTGGCT 6540  
QY 2181 GlyAlaLysLeuProArgTyrGlyAlaLeuArgGlyGluGlnProProAspLeuGluThr 2200  
DB 6541 GGGGCCAAGGTGGCCCGCTACAGAGGCCCTGGGGAGAGAGCCCGGACCTTGAAGACA 6600  
QY 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
DB 6601 AAGGATCTTGGCTGAGTGTGCTTTCAGAGAGACCCCGCTGGTCAAGCCCGAGAGC 6660  
QY 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGluArgArgHisProGluLeu 2240  
DB 6661 CCGGAGAGGCCCAAGAGCAAGAGAGCTGGACGGCGACAGGACGGCACCCGAGAGCTG 6720  
QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
DB 6721 AGCCAGAGGAGGCTGGCCAGCGTCATCATCTACCGACCTTGGCCGGGCTTACCTGCT 6780  
QY 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
DB 6781 CATAACTATGACCTGACCAAGCGCACCTTGAGAGTCCCAAGCCCGATCATCAACA 6840  
QY 2281 ProValValSerTleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLys 2300  
DB 6841 CCGGTGGTAGGATCAGCGTCCATGATGATGAGAGACTTCTGCGCCGGGCGCTGACAA 6900  
QY 2301 ProValThrValGlnPheArgLeuGluThrGluGluArgThrLysProIleCysVal 2320  
DB 6901 CCGGTACGGGTGAGTTCGGCTGCGAGACAGAGAGCGAGACCAAGCCCATCTGTGTC 6960  
QY 2321 PheTrpAsnHisSerTleLeuValSerGlyThrGlyGlyTyrTrpSerAlaArgGlyCysGlu 2340  
DB 6961 TTCTGGAAACATTCAATCCCGTCACTGGCACAGTGGCTGGTGGCCAGAGGCTGTGA 7020  
QY 2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
DB 7021 GTCGTCTTCGCGATAGAGCCACGTCACTGCTCCAGTGCACCAATGACGAGCTTCGCT 7080  
QY 2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuLysThrLeuThr 2380  
DB 7081 GAGCTCATGAGCATTTCTCGCGGAGAAATGGGGAATCTGCTCACTGAAGACATGACA 7140  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePhePheLeuThrLeu 2400  
DB 7141 TACGTGGCTCTAGGTCTCACTGGCTGCCCTTCTGCTCACTCTTCTTCTCACTCTC 7200  
QY 2401 LeuArgTleLeuArgSerAsnGlnHisGlyTleArgArgAsnLeuThrAlaAlaLeuGly 2420  
DB 7201 TTGCGTATCTCGGTGCTCCAAACAGGCACTCCGATCCGATCACTGACCTGGCGC 7260  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyTleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
DB 7261 CTGGCTCAAGTGGTCTTCTCTCGGGAATCAACAGGCTGACCTCCCTTTGGCTGACA 7320  
QY 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGlu 2460  
DB 7321 GTCATTGGCATCTGCTGACACTTCTGACCTTCTGACACTTCTTCTGGGCTCTGCGAG 7380  
QY 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetArg 2480  
DB 7381 GCGTTGACACTGATACCGGACACTCACTGAGTGGCGGATGTCAACACCGGCCCATAGCC 7440  
QY 2481 PheTyrTyrMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
DB 7441 TTCTACTACATGCTGGGCTGGGCGTGGCTGCTTCACTACAGAGGCTTACCGTGGGCTG 7500  
QY 2501 AspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerTleTyrAspThrLeuIle 2520  
DB 7501 GACCCCGAGGGCTACAGGAACTGACTTCTGCTGCTCCATCTATATGACAGGCTCATC 7560  
QY 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrTleLeu 2540

DB 7561 TGGAGTTTGTGGCCCGGCTGGCTTTGGCGTCCGATGAGTGTCTTCTGTACATCTCG 7620  
QY 2541 AlaAlaArgAlaSerCysAlaAlaGlnArgGlnGlyPheGlnLysGlyProValSer 2560  
DB 7621 GGGGCCCGGCTCTGTGTGGCCAGGCGGAGGCTTGTGAAGAAAGAGTCTGTCTCG 7680  
QY 2561 GlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
DB 7681 GGCCTGAGCCCTCTTCCCGCTCTCTGCTGCTGAGGCCACAGTGGTGTGCTGACCTG 7740  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTyrTlePheAlaIleThrCysAsnCysIle 2600  
DB 7741 CTCTGTCAACAGAGACACCTCTCTTCCACTACCTTTGTGATCTGCAATTGCAATC 7800  
QY 2601 GlnGlyProPheIlePheLeuSerTyrValValLeuSerLysGluValArgValAlaLeu 2620  
DB 7801 CAGGCGCCCTTATATCTCTCTCTATGTGTGTAGCGAAGAGTCCGGAACACACTC 7860  
QY 2621 LysLeuAlaCysSerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
DB 7861 AAGCTTGGCTGAGCGCGAGCCAGCCCTGACCTTGTCTGACACAGAGTCCACCTCG 7920  
QY 2641 ThrSerSerTyrAsnCysProSerProTyrAlaAspGlyArgLeuTyrGlnProTyrGly 2660  
DB 7921 ACCGTGCTCAACTGCGCCCGCCCTACAGAGATGGGCGGCTGTACAGGCCCTACGGA 7980  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTyrIle 2680  
DB 7981 GACTGGCGGCTCTCTGACAGACAGTGTGCTGGCGAAGATCAGCCAGCTTACATC 8040  
QY 2681 ProPheLeuLeuArgGluGluSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
DB 8041 CCTTCTTGTGAGGAGAGAGTCCGCACTGAACCTTGCCAGAGGCCCTTGGCTGGGG 8100  
QY 2701 AspProGlySerLeuPheLeuGluGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
DB 8101 GATCCAGGAGCGCTGTCTCTGAGAGTCAAGACAGAGCATGATCTGACACAGCATCC 8160  
QY 2721 AspSerAspLeuSerLeuGluAspAspGlnSerGlySerTyrAlaSerThrHisSerSer 2740  
DB 8161 GACAGTGAACGTCTCTTGAAGACAGACAGAGTGTCTCTTATGCTTACCACTCATCA 8220  
QY 2741 AspSerGluGluGluGluGluGluGluGluGluGluAlaAlaPheProGlyGluGluGly 2760  
DB 8221 GACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGGCGCTTCCCTGAGAGAGAGGCG 8280  
QY 2761 TrpAspSerLeuLeuGlyProGlyAlaGluArgLeuProLeuHisSerThrProLysAsp 2780  
DB 8281 TGGGATACCTGCTGGGCGCTGAGAGAGAGACTGCCCTGCAAGTACTCCCAAGAT 8340  
QY 2781 GlyGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800  
DB 8341 GGGGGCCAGAGGCTGGAGAGGCCCTGGCCAGAGACTTGTGGACCAAGCAAAAGAG 8400  
QY 2801 SerSerGlyAsnGlyAlaProGluGluArgLeuArgGluAsnGlyAspAlaLeuSerArg 2820  
DB 8401 AGTATGGCAACGGGGCCCTGAGAGAGGCGCTGGGAGAAATGAGAGAGCCCTGTCTGA 8460  
QY 2821 GlnGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLysGlyTleLeuLys 2840  
DB 8461 GAGGGGTCTCTAGGCCCCCTTCAAGGCTCTTCTGCTCCAGGCTTCAAAAGGCACTCTTAA 8520  
QY 2841 LysLysCysLeuProThrTleSerGluLysSerSerLeuLeuArgLeuProLeuGluGln 2860  
DB 8521 AAGAGTGTCTGCCACATCAAGAGAGAGAGAGAGAGCTCTCTGCGCTCCCTTGAAGCA 8580  
QY 2861 CysThrGlySerSerArgGlySerSerAlaSerGluGlySerArgGlyGlyProProPro 2880  
DB 8581 TGCAAGAGGTCTTCCCGGGGCTCTCTCGCTAGTAGGAGAGCGGGGCGGCCCCCTTCCC 8640  
QY 2881 ArgProProProArgGlnSerLeuGlnGluGluLeuAsnGlyValMetProIleAlaMet 2900

Db 8641 CGCCACGCGCCCGCAGAGCTTCAGAGCAGCTGAACGGGGTCCATGCCATGCCATG 8700  
 Qy 2901 SerIleValaIaGIYThrValaAspGluAspSerSerGIYpHeuLeuPhePheAsn 2320  
 Db 8701 AGCATCAAGGACGACGATGATGAGACACTGTCAGAGCTCCGAATTCCTCTTAAAC 8760  
 Qy 2921 PheLeuHis 2923  
 Db 8761 TTCCTGCAT 8769

## RESULT 2

US-10-225-567A-523  
 ; Sequence 523, Application US/10225567A  
 ; Publication No. US20030113798A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lifespan Biosciences  
 ; APPLICANT: Brown, Joseph P.  
 ; APPLICANT: Burnet, Glenna C.  
 ; APPLICANT: Roush, Christine L.  
 ; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
 ; FILE REFERENCE: 1920-4-4  
 ; CURRENT APPLICATION NUMBER: US/10/225,567A  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: 60/257,144  
 ; PRIOR FILING DATE: 2000-12-19  
 ; NUMBER OF SEQ ID NOS: 2292  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 523  
 ; LENGTH: 10531  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-225-567A-523

## Alignment Scores:

Pred. No.: 0 Length: 10531  
 Score: 15545.00 Matches: 2923  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 15 Gaps: 0

US-09-916-849a-3 (1-2923) x US-10-225-567A-523 (1-10531)

Qy 1 MetArgSerProAlaThrGIYValProLeuProThrProProProLeuLeuLeu 20  
 Db 63 ATGGAGAGCCGCGGACACGGGCTCCCTCCCAACGGCGCGCGGCTGCTGCTG 122  
 Qy 21 LeuLeuLeuLeuLeuProProProLeuLeuGIYAspGlnValGIYProCYAspSerLeu 40  
 Db 123 TTGCTGCTGCTGCTGCGCGCGGCACTATTGGAGACCAAGTGGGGCTGCTGCTG 182  
 Qy 41 G1SerArgGIYArgGIYSerSerGIYAlaCYAspAlaProMetGIYTrpLeuCYAspSer 60  
 Db 183 GGGTCCAGGGACCAAGGCTCTTCGGGGGCTTGGCCCCCATGGGCTGCTGCTGCT 242  
 Qy 61 SerIaSerAsnLeuTrpLeuTrpThrSerArgCYAspAspAlaGIYThnGluLeuThr 80  
 Db 243 TCAACGTCGAACCTCTGGCTCTACACCAAGCGGTGAGGATGGGGGACAGCTGACT 302  
 Qy 81 G1YHisLeuValProHisIaAspGIYLeuArgValTrpCYAspProGluSerGluHis 100  
 Db 303 GGGCACCCTGTAACCCCAACAGATGGCTGAGGGTGGTGTCCAGATCCGAGGCCCAT 362  
 Qy 101 I1ProLeuProProAlaProGluGIYCYAspTrpSerCYAspGluLeuGIY116G1Y 120  
 Db 363 ATTCCTTACCAACAGCTCTCGAAGGCTGGCCCTGGAGCTGTCCCTCTGAGGATTGGA 422  
 Qy 121 G1YHisLeuSerProGlnGIYLYLeuLeuThrLeuProGluGluHisProCYAspLeuVal 140  
 Db 423 GGGCACCCTTCCCAACAGGCGAAGCTCACTGCCCGAGAGACCCGTTGTTAAAGCT 482  
 Qy 141 ProArgLeuArgCYAspGlnSerCYAspLeuLeuAlaGlnIaProGluLeuArgAlaGIYGlu 160

Db 483 CCAAGGCTCAGATGTCAGTCTCTGACATGCGACAGAGCCCGGCTCAGAGGACGGGAA 542  
 Qy 161 ArgSerProGluGluSerLeuGIYArgArgValArgAsnValAsnThrAlaProGln 180  
 Db 543 AGGTCAACCAAGAGAGCTCTGGGCTGGCGTCCGAAAAGAAATGAATACAGCCCCCAG 602  
 Qy 181 PheGlnProProSerTrpGlnIaThrValProGluAsnGlnProAlaGIYThrProVal 200  
 Db 603 TTCAGCCCCCAAGCTACAGGCGCAAGTCCGGAGAACCAAGCAGACAGCCCTGTT 662  
 Qy 201 AlaSerLeuArgAlaI1aAspProAspGIYGIYValGIYArgLeuGIYTrpMet 220  
 Db 663 GCATCCCTTACAGGCGCTTCCACCGGACGAGGGTGGCGGCTGACTGAGATCACACAG 722  
 Qy 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGIYAla 240  
 Db 723 GATGCCCTCTTGATAGCCGCTCCACAGATCTTCTCTGAGACCAAGTCACTGATGCA 782  
 Qy 241 ValThrThrAlaGluGluLeuAspArgGluThrTrpSerThrHisValPheArgValThr 260  
 Db 783 GTAACCAACGCGAGAGAGCTGGATGCTGAGACCAAGACACCAAGCTTTCAGGGTCAAG 842  
 Qy 261 AlaGlnAspHisGIYMetProArgSerAlaLeuAlaThrLeuThrI1aLeuValThr 280  
 Db 843 GCGCAGAGACCAAGCGCATGCCGACGAGATGCCCTGGCTACACTCACTCTTGTTACT 902  
 Qy 281 AspThrAsnAspHisAspProValPheGluGlnGluGIYTrpLYLeuGluSerLeuArgGlu 300  
 Db 903 GACACCAATGACATACCTGTGTTCAGACAGAGATACAGAGAGCTTCAGGGAG 962  
 Qy 301 AsnLeuGluValGIYTrpGluValLeuThrValArgAlaThrAspGIYAspAlaProPro 320  
 Db 963 AACCTGAGGTTGGCTATAGAGGCTCACTGTCAAGGGCCAGGATGTATGATGCCCTCCC 1022  
 Qy 321 AsnAlaAsnI1aLeuTrpArgLeuLeuGIYSerGIYLeuSerProSerGIYValPhe 340  
 Db 1023 AATGCCAATATTCGTACCGCTGTGAGGGGCTCGGGGCGCCCTCTGAAGCTTT 1082  
 Qy 341 G1uI1aAspProArgSerGIYValI1aArgThrArgGIYProValAspArgGluVal 360  
 Db 1083 GAGATGACCTCTGCTGCTGGGAGTCCAGACCCGTGGCTGTGTGATGGGAAAGGTTG 1142  
 Qy 361 GluSerTrpGlnLeuThrValGluAlaSerAspGlnGIYArgAspProGIYProArgSer 380  
 Db 1143 GAATCTTACAGCTGACGGTACAGGCAAGTGAACAGGGCTGGGACCCCGGCTCCGAGT 1202  
 Qy 381 ThrThrAlaI1aValPheLeuSerValGluAspAsnAspAlaProGlnPheSer 400  
 Db 1203 ACCACAGCGCGTGTCTCTTCTGTGAGAGATGACAAATGATATATCCCCAGTTAAG 1262  
 Qy 401 G1uLYArgTrpValValGlnValArgGluAspValThrProGIYValaProValLeuArg 420  
 Db 1263 GAGAAACGCTATGTGTCACAGTGAAGAGATGTACTCCAGGGGCCCAAGTACCTCCA 1322  
 Qy 421 ValThrAlaSerAspArgAspLYSerAsnAlaValI1aHisTrpSerI1aMetSer 440  
 Db 1323 GTCAAGCCTCGGATGACAAAGGGAGCAATGCGGTGTGCACTATGCAATCAAGAT 1382  
 Qy 441 G1YAsnAlaArgGIYGlnPheTrpLeuAspAlaGlnThrGIYAlaLeuAspValI1aSer 460  
 Db 1383 GGCATAGCTCGGGGACAGTTTATCTGAATGCCCAACACTGAGCTCTGAATGGTGAAG 1442  
 Qy 461 ProLeuAspTrpGluThrThrArgGIYTrpThrLeuArgValArgAlaGlnAspGIY 480  
 Db 1443 CTTCTTGATATGAGACCAAGAGATACCTTACGGGTGAGACACAGATGTGGG 1502  
 Qy 481 ArgProProLeuSerAsnValSerGIYLeuValThrValGlnValLeuAspI1aAsnAsp 500  
 Db 1503 CGTCCCACTCTCTATGTCTCTGCTGTGACAGTACAGGTCTCTGATATCAACAGAC 1562  
 Qy 501 AsnAlaProI1aPheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520  
 Db 1563 AATGCCCAATCTTGTGTACAGACCTTTCAGGCTACTGTCTGAGAGAGCTCCCTTA 1622

QY	521	GLYTYRLEUVALLEUHLISVALGINALILESPRALASPRLAGLYASPARNILALTYGLEU	540
Db	1623	GGSTACSGGTCTCTCCATGTCACAGGCTATGACGGTATGCTGGTGAACAAGCCCGCGCTG	1682
QY	541	GLUTYATAGLEUALAGLYVALGIVHISAPRPHEPPOPHETRIILEASDAANGLYTHRCGLY	560
Db	1683	GAATPACCGCCTTGCTGGGTGGGTGGACATGACTTCCCTTCAACATCAAGACAGGAC	1742
QY	561	TRPILLESERVALALALAGILEUASPARGGLUGLVALASPHEITYSERPHEGLYVAL	580
Db	1743	TGGATCTCTGGGTGGTGGTGAACCTGGACCGGAGGAAAGTTGATTCTACAGCTTTGGGGTGA	1802
QY	581	GLULILAAAPGAPPHISGLYTHTRPROALILEUHPHRLIASERVALASERVALTHTRVAL	600
Db	1803	GAAGCTCGAGACCAATGGCATCTCCAGCATCTGCTCTGGCCAGTGCAGGGTACCTGTGC	1862
QY	601	LEUASPVALAASPAPASNAENPPOTHRPHETHRGINPPOGLUTYTHTRVALARGLEUASN	620
Db	1863	CTGGATGTCACACGACCAACATCCCAACTTTCACCAACGAGTACACAGTGGCTCAT	1922
QY	621	GLUASPRLAALAVAGLYTHSERVALILHTRVALSERVALASPARGAPRALHIS	640
Db	1923	GAGAGTACGCTGTGGGCGCACGAGCTGTGACGGGTGACGCTGTGACCGGAGCTCAT	1982
QY	641	SERVALIIEHTRTYRGINLIEHTRSERGLYASNPTRHARGAASAPGPHESERTIEHTRSER	660
Db	1983	AGTGTCACTACCTTACCAAGATCCACAGTGGCAATACCTGAACCGCTTCTCATCAACGAC	2042
QY	661	GLINSERGJGLYGLYLEUVALSERLEUALALEUPROLEUASPRTYLYLEUGLILARGLIN	680
Db	2043	CAAGTGGTGTGGGGCTGTATCCCTTGCCCTGGCACTGAGACTTCAACTGAGCGGGAG	2102
QY	681	TYRVALLEUALAVAILHTRHISERASPGLYTHTRARGLINASPTRIALAGNILEVALVAL	700
Db	2103	TATGTGTGTGGCTGTACCGCTCCGATGGCACTCGGCGAGACACGCGCACMAATGTGGTGTG	2162
QY	701	ASNVALTHTRAPRALAASNPTRHISARGPROVALPHEGINSERSERHISITYTHTRVALASN	720
Db	2163	AATGTCAACGACCGCCACACCCATCGTCTGTCTTTCAGAGCTCCCACTATACAGTGAAT	2222
QY	721	VALANGLUASPARGPROVALAGLYTHTRHTRVALLEULIESESRALATHRASPGLUASP	740
Db	2223	GTTATATGAGGACCGGCGCGGACGACCAACGGTGGTGATACAGCGCCACCGATGAGGAC	2282
QY	741	THRCGLYGLUASNAIISARGLIEHTRTYRPHENETGLUASPSERTIEPPOGLINPHEARGILE	760
Db	2283	ACAGGTGAGAAATGCCCGCATCACTTACATGAGGACAGCAATCCCCCAATTCGGCATC	2342
QY	761	ASPRILASPTRARGLYALAVAILHTRHGINALAGILEUASPRTYRGLUASPGLIVALISER	780
Db	2343	GATCGACGACACGGGGGCTGTACCAACGAGCTAGCTGAGACTACGAGACCAAGTGTCT	2402
QY	781	TYRTHRLEUALAILETHRILASARGASPANGIYLIEPPOGLINYSERASPTRHTRTYR	800
Db	2403	TACACCCCTGGCCATTAAGTCTCGGGACATGGCAATTCGCCAGAGTCCGACCAACCTTAC	2462
QY	801	LEUGLILIELEUVALASNPVALIASNPASNAENLABPPOGLINPHELEUARGASPSEITYR	820
Db	2463	CTGGAGATCTCGGTGAAGACGATGAATACATGCGCTCCAGTTCCTGCGAGACTCTTAC	2522
QY	821	GLINGLYSERVALTYRGLUASPVALIIPROPHETHRHTRSERVALILEUGNILESERVALTHR	840
Db	2523	CAGGGCAGGTGCTATGAGGAGTGTGCACCTTCACTAGCGTCTCGAGATCTAGCCACT	2582
QY	841	ASPRARGASPSERGYLEUANGILYARGVALPHEITYRTHRPHEINGJGLYASPAPGLY	860
Db	2583	GATGTGTATCTCGACCTTAATGGACGGGTCTTCTTACACTTCCAAAGGAGGCGAAGATGGA	2642
QY	861	ASPRGLYASPRHEILIEVALGLUSERTHSERGJLYILEVALARGTHRLEUARGLEUASP	880
Db	2643	GACCGTACCTTATTTGTTGAGTCCAGCTCAAGCAATCGTGCAAGCAAGCTACGAGGCTGAT	2702

QY	881	ArgGIuSenValIaGIuTYrValIeuAAlaTYrValaValaAspIySGIyMeProPro	900
Db	2703	CGAGAGAACGTGGCCCGCATATGCTTCGGGCATATCACTGGACAGAGGGATCCCCCA	2762
QY	901	AlaArgThrProMetGIuValIThrValIThrValIeuAspValaAspAsmProProVal	920
Db	2763	GCCCGCACACCTATGGAAGTGAACGTACACTGTGTGTGATGTGAATGACAAATCCCCCTGTC	2822
QY	921	PheGIuGIuAspGIuPheAspValaPheValGIuGIuAspSerProIIeGIyLeuAlaVal	940
Db	2823	TTTACGACGGATGATGTTGATATGTTGTGTGAGAGAGAACGCCCATTTGGGCTGACCGGTG	2882
QY	941	AlaArgValIThrAlaThrAspProAspGIuGIyThzAsnAlaGIuIIeMetYrGIuIIe	960
Db	2883	GCCCGGGGTCAAGCCACTGACCCCGATGAAAGGACCAATCCCAAGATTATATGACAAATT	2942
QY	961	ValGIuGIyAsnIIeProGIuValaPheGIuLeuAspIIePheSerGIyGIuLeuThrAla	980
Db	2943	GTGAGAGGCAACATCCCTGAGGCTTCCAGCTGGACATCTTCTCCGGGAGCTGACAGCC	3002
QY	981	LeuValaAspLeuAspYrGIuAspAspProGIuTYrValIeuValIleGIuAlaThrSer	1000
Db	3003	CTGTGTAACTTAACTACAGAGACCGGCTGAGTACGTCTGTGTATCCAGGCCAGCTCA	3062
QY	1001	AlaProLeuValSerAArgAlaThrValaIAsValaArgLeuLeuAspAAspAspPro	1020
Db	3063	GCTCTCTGTGTAGCCCGGCTCAAGTCAAGTCCGCTCTTGAACCGCATATGACCAACCA	3122
QY	1021	ProValIeuGIyAsnPheGIuIIeLeuPheAsnAspTYrValIThrAsnAArgSerSer	1040
Db	3123	CGAGTGTGGGCAACTTGAAGATCTTATTCAACAACATATATCAACAAATCCCTCAAGACG	3182
QY	1041	PheProGIyGIyAlaIleGIyAArgValaProAlaIAsAspProAspIIeSerAspSerLeu	1060
Db	3183	TTCCCTGGGGGTGGCAATGGCCGAGTACCTGGCCATGACCTGATTCTCAGATAGTCTG	3242
QY	1061	ThrTYrSerPheGIuAspGIyAsnGIuLeuSerLeuValIleuLeuAsnAlaSerThGIy	1080
Db	3243	ACTTACAGCTTTAGCGGGGAAATGAATCAAGCTGTGCTCTGCAATGTCTCCACGGGT	3302
QY	1081	GIuLeuValSerLeuSerAArgAlaLeuAspAsnAspArgProLeuGIuAlaIleMetSerVal	1100
Db	3303	GAGCTGAAGCTAAGCCCGCATCTGGACACAAACCGGCTCTGGAGGCCATCATAGGCTG	3362
QY	1101	LeuValSerAspGIyValaIAsSerValIThrAlaGIuCYaAlaLeuAArgValIThrIleIe	1120
Db	3363	CTGTGTTCACACGGCGGTACACACCGTGAACCGCCACAGTCCGCTGTGACCATCATC	3422
QY	1121	ThrAspGIuMetLeuThrIAsSerIIeThrLeuAArgGIuAspMetSerProGIuAArg	1140
Db	3423	ACCGATAGATGTCAACCCACAGCATACCGCTGGCTGGAGGACATGTACCCGAGCGC	3482
QY	1141	PheLeuSerProLeuLeuGIyLeuPheIleGIuAlaValaIAlaIaThrLeuAlaThrPro	1160
Db	3483	TTCTCTGCACACATGCTAGGCGCTTCTTCATCCAGGCGGTGGCCCGCACGCTGGCACGCA	3542
QY	1161	ProAspPheIaValaValaPheAsnValaGIuAArgAspThrAspAlaProGIyGIyAsIleIe	1180
Db	3543	CCGACACACGTGTGTGCTTCAACGTACAGCGGAGACACGACGCCCCCGGGGGGCAATC	3602
QY	1181	LeuAsnValSerLeuSerValaGIyGIuProProGIyProGIyGIyGIyProPheLeu	1200
Db	3603	CTCAACTGAGCTGTGTGTGTGGGCGACCGCCAGGGCGCGGGGGCGGCGCTTCTCTG	3662
QY	1201	ProSerGIuAspLeuGIuArgGIuAArgLeuTYrLeuAsnAArgSerLeuLeuThrAlaIleSer	1220
Db	3663	CCCTCTGAGGACCTGCGAGGCGCGCTTATACCTCAACCGCACGCTCTGACGCGGCATCTCG	3722
QY	1221	AlaGIuAArgValaLeuProPheAspAspAsnIIeCYeLeuAArgGIuProCYySGIyAsnTYr	1240
Db	3723	GCACAGCGCTGTGCTCCCTTCCAGACGACACATCTGCTGTGGGAGCCCTTCGAGAACTAC	3782
QY	1241	MetArgCYaValSerValLeuAArgPheAspSerSerAlaProPheIleAlaSerSerSer	1260

Db	3783	ATGGCGCTCGGTGCTGGTGGCTTGCACCTCTCCGGCCCTTCAATGGCTCTCTCC	3842
QY	1261	VALLLeuPheAaPProILLeHIAProValGlyLeuArgCyBaRGCyAProProGlyIyPhe	1280
Db	3843	GGCTCTTCCCGCCCATCCACCCCTCGCAAGGCTGGCTCGCTCGCTGCCCGCCGCTTC	3902
QY	1281	ThRGlyAaPryCySGlyuThRGlyuValAAsPLeuCySryTSerAaRGProCySGlyIProHIs	1300
Db	3903	ACGGGTAACTACTGGAGAACCGAGGTGAACCTCTCTACTGTGGGCTCTGTGGCCCCAC	3962
QY	1301	GlyAaRGCyBaRGSerAaRGlyuGlyIyTyTyThRGyALeuCyBaRGAspGlyTyTyThRGly	1320
Db	3963	GGGCGCTCGCCGACCGCGAGGGCGGCTCAACCTCGCTCTGCGTGAAGGCTTACACGGGT	4022
QY	1321	GIuHIsCySGlyuValSerIaAaRGSerGlyAaRGCySThRGProGlyValCyAlyAaSGly	1340
Db	4023	GAGCACTGTGAAGTAAGTCTGTGCTCAAGCGCTTGCACCCCGGGGTGTGCAGAAATGG	4082
QY	1341	GlyThRCySValAAsmLeuLeuValGlyGlyPheLyCyAaSPCyAProSerGlyAaSPhe	1360
Db	4083	GGCACTGTGTCAACTCTGTGTGGGGGTTCAAGTGCATGGATGCCATCTGGAGACTTC	4142
QY	1361	GIuLyAProTyCySGInValThThRAaRGSerPheProAlaHIsSerPheILeThPhe	1380
Db	4143	GAGAAAGCCCTACTGCAAGTGAACACGGCGACTTCCCGGCCACTCTCTCATCACTT	4202
QY	1381	ArgGlyLeuAaRGInAaRGPheHIsPheThrLeuAlaAaSerPheAlaThThyGlyAaRG	1400
Db	4203	CGCGGCTCGCCGACCGCTTCCACTTCACTTCACTCGCCCTCTGTTTGCACAAAGAGCGC	4262
QY	1401	AaPGlyLeuLeuLeuTyTyAaSGlyAaRGPhaAsnGlyAaRGPhaAsnGlySHIsAaSPheValAlaLeuGly	1420
Db	4263	GACGGGTGTGCTGTATCAATGGCGGTTTCATGAAGAAGCATGACTTGTGGCCCTCGAG	4322
QY	1421	VallIeGInGlyuValGInLeuThRPheserIaGlyGlyuSerThThThValSer	1440
Db	4323	GTGAATCCAGAACAGGTCCAGTCACTTCTGTGAGGAGATCAACCAACGAGTGC	4382
QY	1441	ProPheValProGlyGlyValSerAaPGlyGInTPHIsThRValGInLeuTyTyTy	1460
Db	4383	CCATTCTGTGCCGGAGAGATCAGTATGGCCAGTGGCATTAAGGTGACGTGAATAATCAC	4442
QY	1461	AaTyTyAProLeuLeuGlyGlyInThRGlyLeuProGInGlyProSerGlyuGlyuValAla	1480
Db	4443	AATAAGCCACTGTGTGGTGTCAAGAGGGCTTCCACAGGGCCCATCAAGAACAAAGTGGCT	4502
QY	1481	ValValThRValAaPGlyCyAaSPThRGlyValAlaLeuAaGPheGlySerValLeuGly	1500
Db	4503	GTGGTGAACCGTGAAGCTGTGAACAGAGATGGCTTGGCTTGCATCTGTCTGGGC	4562
QY	1501	AaTyTySerCyAaAlaGInGlyThRGInGlySerTyAlySerTyAaSPLeuThR	1520
Db	4563	AACTACTCTGTGCTGCCACAGGGCACCAAGGTGTGCACAAAGAGTCTGTGAATCTGACG	4622
QY	1521	GlyAProLeuLeuLeuGlyGlyValProAaSPLeuProGlyuSerPheProValAaRGMetAaRG	1540
Db	4623	GGGCCCCGTGCTAATGAGCGGGGTGCTTAACCTGGCCGAGGCTTCCAGTCCGAATGGG	4682
QY	1541	GInPheValGlyCyAaMeAaRGaAsnLeuGInValAaSPSerATySHIsAlaSPMetAlaAaP	1560
Db	4683	CAGTTCGTGGGCTGTGATCGGAGCACTGTCAAGGTGACACCGGCACTAGACATGGCTGAC	4742
QY	1561	PheILeAlaAaSPHAsnGlyThTyValProGlyCyAProAlaTyAlyAaSPValCyAaSPSer	1580
Db	4743	TTCATTGCCAAATAGGACCGGTGCTGTGGCTCTGTGCACAAAGAAAGCTGTGTGACAGC	4802
QY	1581	AaThThRCySHIsAaSGlyGlyTyThRCyValAaSGInTPAaPAlaPheSerCyGlyuCyS	1600
Db	4803	AACAATTGCCAAATAGGGGGCACTTGGCGTGAACAGTGGAGACGGTTCAAGCTGGAGATGC	4862
QY	1601	ProLeuGlyPheGlyGlyTyAaSerCyAaAlaGInGlyMetAlaAaSPProGInHIsPheLeu	1620

Dd	4863	CCCCCTGGGGCTTTGGGGGGCAAGAGCTGGGCCCAAGAAATGGCCAAATCCAGACACTTCCTG	4922
Qy	1621	GIysSerSerLeuValAlaTPPHiEGLYLeuSerLeuProIleSerGlnProTPYLeu	1640
Dd	4923	GGCGAGGAGCCCTGGGGCTTGCGATGGCTCTCCCTCGCTCCCACTCCCAACCTGGTACTTC	4982
Qy	1641	SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly	1660
Dd	4983	AGCTCAATGTTCCGACAGCCGCCAGCCGACAGGTGTCTGCTGACGGGCATCACAGGGGG	5042
Qy	1661	ArgSerThrIleThrLeuGlnLeuArgGlnGlyYhiIvaIleMetLeuSerValGIuGlyYThr	1680
Dd	5043	CGCAGCACCAATCACCCCTAACAGCTAACGAGAGGGGCACACTGATGCTGAGCCGAGGGGCACA	5102
Qy	1681	GIYLeuGlnAlaSerSerLeuArgLeuGlnProGIYArgAlaAspAspGlyAspTPPHiS	1700
Dd	5103	GGGCTTAAAGGCTCTCTCTCTCCGTCTGAGACAGAGCCGGGCCAATGACGGTACTGGCAC	5162
Qy	1701	HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGIYhiIaIleLeuSerPheAsp	1720
Dd	5163	CATCAACAGCTGGGCACTGGGAGCCAGGGGGGGCTGGCCATGCATTCATTCGTCCTTGAT	5222
Qy	1721	TYrGIYglnGlnIleArgAlaGlnGlyYAsnLeuGlyProArgLeuHiEGLYLeuHiSLeuSer	1740
Dd	5223	TYATGGGACGAGAGAGAGAGAGGGCCAACTGGGGCCCCCGGCTGCATGTCTGCACCTGAGC	5282
Qy	1741	AsnIleThrValGIYGIYIleProGIYProAlaGIYAlaAlaArgGlyPheArgGly	1760
Dd	5283	AACATAACAGTGGGGCGGAAATACCTGGGGCCAGCCGGCGGTGTGGCCCGTGGCTTTCCGGGC	5342
Qy	1761	CysLeuGlnGlnIYAlaArgValSerAspThrProGlnGlyYValAsnSerLeuAspProSer	1780
Dd	5343	TGTTTGACAGGGGTGTGGCGGTGACCGATACCCACAGAGGGGGGTTPAACAGCTCGATATCCAGC	5402
Qy	1781	HisGIYGIuSerIleAsnValGIuGlnGlyCysSerLeuProAspProCysAspSerAsn	1800
Dd	5403	CATGGGAGAGACATCAACGTGAGAGAGAGGCTGTAGACTGCTGACCTTTGTACTCAAC	5462
Qy	1801	ProCysAspProAlaAsnSerTYrCysSerAspAspTPHAspSerTYrSerCysSerCysAsp	1820
Dd	5463	CCGGTCTCTGCTTAACAGCTATTCCAGCAAGACAGCTGGGACAGCTATTCTTCACACTGTGAT	5522
Qy	1821	ProGlyTYrTYrGlyYAspAsnCysThrAsnValCysAspLeuAsnProCysGlnIleGln	1840
Dd	5523	CGAGGTACTAGGTGTGACACTGTACTAAATGTGTGTGACTTGAAACCCGTGTAGACACAG	5582
Qy	1841	SerValCysThrArgIysProSerAlaProHiEGLYrThrCysGlnCysProProAsn	1860
Dd	5583	TCGTGTGTATCCGGAAAGCCCAAGTGGCCCCCAATGGGTATACGTGAGGTGTCCCCCAAT	5642
Qy	1861	TYrLeuGlyProTYrCysGlnIleThrArgIleAspGlnProCysAspProArgIleTYrTPHGIY	1880
Dd	5643	TACCTTGGGCAATACGTGTAGAACAGAGATTGACACAGCTTGTCCCCGTGGCTGGTGGGA	5702
Qy	1881	HisProThrCysGlyProCysAsnCysAspValSerIysGlyPheAspProAspCysAsn	1900
Dd	5703	CATCCCACTATGGGCCCAATGACACTGTGACTGTACGAAAGCTTTGACCCAGACTGAC	5762
Qy	1901	LYrThrSerGlyGlnCysHiEGYLYYLeuGlnAsnHiSTYrArgProProGIYSerProThr	1920
Dd	5763	AAGCAAGAGCGGCGAGTGCACCTGCAGAGAAACACTACCGGGCCCCCAGGCAAGCCCCACC	5822
Qy	1921	CysLeuLeuCysAspCysTYrProThrGlySerLeuSerArgValCysAspProGluAsp	1940
Dd	5823	TGCTCTTGTGTACTCTACCCCAAGGCTCTTGTGCCAGATCTGTACCCCTGAGAT	5882
Qy	1941	GIYGIuCysProCysLYsLYsProGIYValIleGIYArgGlnCysAspArgCysAspAsnPro	1960
Dd	5883	GGCGAGGTGCATGCAAGCAGAGGTGTCAATGAGGCGTCAAGTGAACCGCTGTGACACCT	5942
Qy	1961	PheAlaGluValThrThrAsnGlyCysGlnValAsnTYrAspSerCysProArgAlaIle	1980
Dd	5943	TTTGTCTAGGTCAACCAATAGGTGTGAAGTAAATTATAGACAGCTGCCCAAGAGCAATT	6002

QY 1981 GlnAlaGlyTlLeTrrPrrProAArgThrArgPheGlyLeuProAlaAlaAlaProCyPro 2000  
DB 6003 GAGGCTGGGATCTGGTGGCCCGGTAACCGGCTTGCGGCTGCTGCTCCCTGACCC 6062  
QY 2001 LysGlySerPheGlyThrAlaValArgHisCyAspProLinhSAArgGlyTrrPLeuProPro 2020  
DB 6063 AAGAGCTCTTGTGGAGCTGCTGAGCGCACTGTATAGACACAGGGGGGTGGCTCCCGCCA 6122  
QY 2021 AsnLeuPheAsnCyThrSerLleThrPheSerGlyLeuGlyPheAlaGlnArgLeu 2040  
DB 6123 AACCTCTTCAAGTCAAGTCACTTCACTTCAAGAACTGAAGGGCTTCGCTGAGCGGCTA 6182  
QY 2041 GlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeuAlaLeuLeuArg 2060  
DB 6183 CAGGGGATAGTCAAGGCTTGAAGTCAAGGCGCTCCACAGCAGTACCGCTGCTCTGCGC 6242  
QY 2061 AsnAlaThrGlnHisThrAlaGlyTrrPheGlySerAspValLysValAlaTrrGlnLeu 2080  
DB 6243 AACGCCACGGACACACAGCTGCTGCTACTTGGCGACACCTCAAGGTGGCTTACCAAGCTG 6302  
QY 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyLeuSerAlaThrGln 2100  
DB 6303 GCCACGGGGTGTGGCTGGCCACAGAGACACCAAGGGGGCTTGGGCTGTCTGCCACACAG 6362  
QY 2101 AspValHisPheThrGlnAsnLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
DB 6363 GAGGTGCACTTCACTGAGAACTGCTGCGGGTGGCAGCGCCCTCTGACACAGCCAC 6422  
QY 2121 LysArgHisThrGlnLeuLleGlnGlnThrGlnGlyTrrAlaTrrPLeuLeuGlnHis 2140  
DB 6423 AAGGGGACATGGAGCTGATTCAGACACAGAGGGTGGCAACCGCTGGCTGCTCCAGCAC 6482  
QY 2141 TyrGlnAlaTrrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTrrLeuSerProPhe 2160  
DB 6483 TATGAGGCTTACGCGAGTGGCTTGGCCCAAGACATGGGGCACACTTACCAAGCCCTTC 6542  
QY 2161 ThrTlLeuAlaThrProAsnLleValLleSerValValArgLeuAspLysGlyAsnPheAla 2180  
DB 6543 ACCATCTCTACGCGCCCAACATTTGATCTCCGTAGTGGCTTGGACAAAGGAACTTTGGCT 6602  
QY 2181 GlnAlaLysLeuProArgTrrGlnAlaLeuArgGlyGlnGlnProProAspLeuGlnThr 2200  
DB 6603 GGGGCGCAAGTGGCCCGCTACAGAGGCTCTGTGGAGAGACGCCCGCGACCTTGAGACA 6662  
QY 2201 ThrValLleLeuProGlnSerValPheArgGlnThrProProValValArgProAlaGly 2220  
DB 6663 ACAGTCACTTGTGCTGAGTGTGTCTTCAAGAGACGCCCCCGTGTCAAGGCCCGCAGGC 6722  
QY 2221 ProGlyGlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgArgHisProGlnLeu 2240  
DB 6723 CCGCGAGAGGCCACAGGACAGAGAGTGGCAAGGCAAGCAAGCAAGCCGAGAGCTG 6782  
QY 2241 SerGlnGlyGlnAlaValAlaAsnValLleLleTrrArgThrLeuAlaGlyLeuLeuPro 2260  
DB 6783 AGCCAGGGGTAGGCTGTGGCCAGCGTATCATTCACGACACCTTGGCCGGGCTTACGCTC 6842  
QY 2261 HisAsnTrrAspProAspLysArgSerLeuArgValProLysArgProLleLleAsnThr 2280  
DB 6843 CATAACTATGACCTTGAACAGCCAGCTTGAAGTCCCAAGGCCCGATCAACACACA 6902  
QY 2281 ProValValSerLleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspLys 2300  
DB 6903 CCGGTGTGAGCATCGCTGCATGATGAGAGAGGTCTTGGCCCGGGCCCTGAGCAAA 6962  
QY 2301 ProValThrValGlnPheArgLeuLeuGlnThrGlnGlnArgThrLysProLleCyVal 2320  
DB 6963 CCGGTACGGGTGAGTTCGGCTGTGGAGACAGAGGAGGAGCAAGCCCATCTGTGTC 7022  
QY 2321 PheTrrAsnHisSerLleLeuValSerGlyTrrGlyTrrSerAlaArgGlyCyGln 2340  
DB 7023 TTCTGGAAACATTCAATCTCTGTCAGTGGCAAGGTGGCTGTGCTGGCCAGAGGCTGTGA 7082

QY 2341 ValValPheArgAsnGlnSerHisValSerCyGlnCyAsnHisMetThrSerPheAla 2360  
DB 7083 GTGCTCTTCCGACATGAGAGCCACGTCACTGCAATGGACACCATGACAGCTTGCT 7142  
QY 2361 ValLeuMetAspValSerArgArgGlnGlnGlnLleLeuProLeuLysThrLeuThr 2380  
DB 7143 GTGCTATGAGACCTTCTCTGGCGGAGAAATGGGAGATCTGTCACTGAGACATGACA 7202  
QY 2381 TyrValAlaLeuGlyValThrLeuAlaLeuLeuLeuThrPhePheLeuThrLeu 2400  
DB 7203 TACGTGGCTTGAAGTTCACCTTGGCTGCTTCTGTCACTTCTTCTTCTTCACTTC 7262  
QY 2401 LeuArgLleLeuArgSerAsnGlnHisGlyLleArgArgAsnLeuThrAlaAlaGly 2420  
DB 7263 TTGCGATCTGTGGCTTCAACCAACAGCGCATCCAGTACCTGACAGTGGCCGTGGGC 7222  
QY 2421 LeuAlaGlnLeuValPheLeuLeuGlyLleAsnGlnAlaAspLeuProPheAlaCyThr 2440  
DB 7323 CTGGCTCAGCTGTCTTCTCTGGGAAATCAACAGGCTGACCTCTTGTGGCTGACCA 7382  
QY 2441 ValLleAlaLleLeuLeuHisPheLeuTrrLeuCyThrPheSerTrrAlaLeuLeuGln 2460  
DB 7383 GTCAATTGCCATCTGTGACCTTCTGTACCTTGTCACTTCTTCTGCGCTGCTGGAG 7442  
QY 2461 AlaLeuHisLeuTrrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArg 2480  
DB 7443 GCCTTCACTGTACCGGGCACTCACTGAGGTGGCGGATGTCAACACCGGCCCCATGGCG 7502  
QY 2481 PheTrrTrrMetLeuGlyTrrGlyValProAlaPheLleThrGlyLeuAlaValGlyLeu 2500  
DB 7503 TTCTACTACATGCTGGGGCTGGGGCGTGGCTGCTTATACAGGGCTTACCGGGGCTG 7562  
QY 2501 AspProGlnGlyTrrGlyAsnProAspPheCyStrPLeuSerLleTrrAspThrLeuLle 2520  
DB 7563 GACCCGAGGGCTACGGGAAACCTGACTTCTGTGCTTCTTCACTTATGACAGCTGATC 7622  
QY 2521 TrrSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTrrLleLeu 2540  
DB 7623 TGGAGTTTGTGGCGCCGGGTGGCCTTTGGCGTGTGATGAGTGTCTTCTGTATCATCTG 7682  
QY 2541 AlaAlaArgAlaSerCyValAlaAlaGlnArgGlnGlyPheGlnLysGlyProValSer 2560  
DB 7683 GCGGCCCGGCTCTCTGCTGCTGCCAGCGGCGGCTTGGAAAGAAAGATCTGTCTCG 7742  
QY 2561 GlnLeuGlnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrrPLeuLeuAlaLeu 2580  
DB 7743 GGCCTTCAAGCTCTTCTGCGCTCTTCTGTGCTGAGGCGCACAGTGGCTGTGGACTG 7802  
QY 2581 LeuSerValAsnSerAspThrLeuLeuPheHisTrrLeuPheAlaThrCyAsnCyHis 2600  
DB 7803 CTCTGTGTCAAGGACACCTCTCTTCACTACCTTGTGTATCTGCAATTGCATTC 7862  
QY 2601 GlnGlyProPheLlePheLeuSerTrrValValLeuSerLysGlnValArgLysAlaLeu 2620  
DB 7863 CAGGGCCCTTATATCTTCTCTTCAATGTGGCTTGAAGAGAGGTGGAGAAAGACTTC 7922  
QY 2621 LysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
DB 7923 AAGCTTGGCTGAGCGCGAGCCAGCCAGCTTACCTGTCTTACACCAAGTCCACCTG 7982  
QY 2641 ThrSerSerTrrAsnCyAspProSerProTrrAlaAspGlyArgLeuTrrGlnProTrrGly 2660  
DB 7983 ACCGTGCTCTTACACTGGCCCGCCCTTACAGAGTATGGCGGCTGTACACAGCCCTAAGGA 8042  
QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTrrLle 2680  
DB 8043 GACTGGCCGGCTCTCTGTGACAGACCAAGTGTGGGAGAAAGTCAAGCCATCACTATC 8102  
QY 2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
DB 8103 CCTTCTTGTGAGGAGAGAGTCCGCACTGAAACCTTGGCCAAAGGCCCTTGGCTGGGG 8162  
QY 2701 AspProGlySerLeuPheLeuGlnGlyLysArgGlnGlnHisAspProAspThrAspSer 2720



```

Db      8163 GATCCAGGACGCTGCTTCTGAGAGTCAAGACGAGCATGATCTGACACGACTCC      8222
Qy      2721 AapSerAapLeuSerLeuGluAapAapGluSerGlySerYrAlaSerThrHisSerSer 2740
Db      8223 GACAGTGAACCTGCTCTTGAAGACACACAGATGCTCTTATGCTCTTACCACTCACTCA 8282
Qy      2741 AapSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 2760
Db      8283 GACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8342
Qy      2761 TrpAapSerLeuLeuGlyProGlyValAgluAargLeuProLeuHisSerThrProlyAap 2780
Db      8343 TGGGATGACCTGCTGAGGCTGAGACACAGAGACTGCCCCGACAGATCTCCCAAGAT 8402
Qy      2781 G1GlyIProGlyIProGlyIYrAlaProIrrProGlyAapPheGlyYrThrAlaYrGlu 2800
Db      8403 GGGGGCCCAAGGGCCCTGGCAAGGCCCCCTGGCCAGAGACTTTGGACACACAGCAAAAGAG 8462
Qy      2801 SerSerGlyAamGlyAlaProGluGluAargLeuAargGlyAamGlyAapAlaLeuSerAarg 2820
Db      8463 AGTAGTGGCAACGGGGCCCTGTAGAGAGCGGCTGCGGAGAAATGAGATGCCCTGTCTCGA 8522
Qy      2821 GluGlySerLeuGlyIProLeuProGlyYrSerSerAlaGlnProHisGlyGlyIleLeuLys 2840
Db      8523 GAGGGGTCTCTAGGCCCCCTTCCAGGCTCTTCTGCCAGGCTCAACAAAGGACATCTTAAAG 8582
Qy      2841 LysLysCybleuProThrIleSerGluYrSerSerLeuLeuAargLeuProLeuGluGlu 2860
Db      8583 AAGAAGTCTCTGCCACATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8642
Qy      2861 CysThrGlySerSerAargGlyYrSerSerAlaSerGluGlyYrSerAargGlyIleProPro 2880
Db      8643 TGCAAGAGGCTCTTCCCGGGGCTCTCTCCGCTAGTAGAGGAGAGAGAGAGAGAGAGAGAGAG 8702
Qy      2881 ArgProProProAargGluSerLeuGluGluGluGluGluGluGluGluGluGluGluGlu 2900
Db      8703 GCGCCACGCGCCCGGACAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8762
Qy      2901 SerIleYrAlaGlyYrThrValAapGluAapSerSerGlySerGluPheLeuPheAasn 2920
Db      8763 AGCATTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8822
Qy      2921 PheLeuHis 2923
Db      8823 TTCTCGCAT 8831

```

## RESULT 3

```

US-09-788-711A-1
; Sequence 1, Application US/09788711A
; Patent No. US20020058328A1
; GENERAL INFORMATION:
; APPLICANT: Tania Tamsin Testa
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30225
; CURRENT APPLICATION NUMBER: US/09/788,711A
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 0004196.2
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 8871
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-788-711A-1

```

## Alignment Scores:

```

Pred. No.: 0 Length: 8871
Score: 15518.50 Matches: 2923
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 0
Query Match: 99.83% Indels: 33

```

```

DB: 9 Gaps: 1
US-09-916-849A-3 (1-2923) x US-09-788-711A-1 (1-8871)
Qy      1 MetLysSerProAlaThrGlyValProLeuProThrProProProProLeuLeuLeu 20
Db      1 ATGCGAGAGCCCGGACACGGGCTCCCTCCCAACGGCCGCGCGCTGCTGCTCTG 60
Qy      21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAapGlnValGlyProCyAargSerLeu 40
Db      61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qy      41 GlySerAargGlyAargGlyYrSerSerGlyValAaAlaProMetGlyYrPLeuCyProSer 60
Db      121 GGGTCCAGAGGAGACAGAGGCTCTTGGGGGGCTGCGGCCCTCAATGGGCTGGCTGTGCATCC 180
Qy      61 SerLaserLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
Db      181 TCAGCTGCAACCTCTGCTCTTACACAGCGGCTGCGAGGAGATGCGGACACTGAGCTACT 240
Qy      81 GlyHisLeuValProHisLaserGlyLeuAargValTrpCyProGluSerGluAlaHis 100
Db      241 GGGCACCCTGTAACCCCAACAGATGGCTGAGGGTTGGTGTCTCAAAATCCAGAGGCCAT 300
Qy      101 IleProLeuProProAlaProGluGlyCyAproIrrSerCyAargLeuLeuGlyIleGly 120
Db      301 ATTCCTCTACCAACAGCTCTTGAAGGCTGCCCTGAGAGCTGTGCTCTCTGAGCATTTGA 360
Qy      121 GlyHisLeuSerProGluGluYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYrYr 140
Db      361 GGGCACCCTTCCCAACAGGGGCAAGCTCACAGCTGCCCAAGAGAGAGAGAGAGAGAGAGAG 420
Qy      141 ProAargLeuAargCyGlnSerCybleuAlaGlnAlaProGlyLeuAargAlaGlyGlu 160
Db      421 CCAAGGCTCAGATGCCAGTCTGCAAGCTGCAAGGCCCCGGGCTCAAGGCAAGGAGAA 480
Qy      161 AargSerProGluGluSerLeuGlyYrAargGlyYrAargAanValAanThrAlaProGln 180
Db      481 AGGTCAACAGAAAGTCTCTGGGGGCTGCGGAGAAAGATGAAATTAACAGCCCCCAG 540
Qy      181 PheGlnProProSerYrGlnAlaThrValProGluAamGlnProAlaGlyYrProVal 200
Db      541 TTCCAGCCCCCACTACAGGCCACAGTCCGAGAAACAGCCAGAGAGAGAGAGAGAGAGAG 600
Qy      201 AlaSerLeuAargAlaLaserProAapGluGlyIleValAglAargLeuGluYrYrMet 220
Db      601 GCATCCCTGAGGGGCAATGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      221 AapAlaLeuPheAapSerAargSerAargGlnPhePheSerLeuAapProValThrGlyAla 240
Db      661 GATGCCCTCTTGTATGATCCGCTCCACAGATTTCTTCCCTGAGCCACAGTCACTGGTGA 720
Qy      241 ValThrThrAlaGluGluLeuAapAargGluThrYrYrSerThrHisValPheAargValThr 260
Db      721 GTTACCAACAGCCAGAGAGAGAGAGATGTAAGACAAAGAGACACCAAGTCTTCAAGGTACG 780
Qy      261 AlaGlnAapHisGlyMetProAargSerAlaLeuAlaThrLeuThrIleLeuValThr 280
Db      781 GCGCAGAGACACAGGCAATGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy      281 AapThrAapAapHisAapProValPheGluGluGluGluGluGluGluGluGluGluGlu 300
Db      841 GACACCAATGACCAATGACCTGTGTTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy      301 AasnLeuGluValGlyYrGluValLeuThrValAargAlaThrAapGlyAapAlaProPro 320
Db      901 AACCTGAGAGTGTGCTATGAGGTCTCACTGTCAAGGCCACAGATGTATGATGCCCTCCC 960
Qy      321 AasnAlaAanIleLeuYrAargLeuLeuGluGlyYrSerGlyYrSerProSerGlyValPhe 340
Db      961 AATGCCAATATTTCTGTATCCCTGCTGAGAGGGGTCTGGGGGAGCCCTCTGAAATCTTT 1020
Qy      341 GluIleAapProAargSerGlyValIleAargThrAargIyProValAapAargGluGluVal 360

```



```

Db 1021 GAGATCGACCCCTCGCTGGGGTGAATCCGAACCCCTGGCCCTGGATCGGGAGAGAGTGT 1080
Qy GluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArgSer 380
Db 1081 GAATCTTCACAGCTGACGGTGAAGGACCAAGTACAGGGGTCCGAGACCCGGGTCTCCGAGT 1140
Qy ThrThrAlaAlaValPheLeuSerValGluAspAspAsnAlaProGlnPheSer 400
Db 1141 ACCACAGCCGCTGTTTCTTCTGTGTGAGGATGACATGATTAATGCCCAAGTTAAT 1200
Qy GluGlySerGlyTyrValGlnValArgGluAspValThrProGlyAlaProValLeuArg 420
Db 1201 GAGAGGCGCTATGTGTCCAGGTGAGGAGATGATGATCCAGGGGCCCTGACTCCGA 1260
Qy ValThrAlaSerAspArgAspGlySerAsnAlaValAlaHisTyrSerIleMetSer 440
Db 1261 GTCAACGCTCGGATTCAGACCAAGGGAGACATGCCGTGGTGCATTAAGCATGAGT 1320
Qy GluAlaAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValValSer 460
Db 1321 GGCAATGCTCGGGACAGTTTATCTGGATGCCCAAGCTGGAGCTCGGATGTGGTGGC 1380
Qy ProLeuAspTyrGlnThrThrTyrGlyGlnTyrThrLeuArgValArgAlaGlnAspGly 480
Db 1381 CCTTTGACTATGAGACGACCAAGAGATGACCCCTACGGGTGCGACGACAGATGTGGC 1440
Qy ArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsnAsp 500
Db 1441 CGTCCCACTCTTAATGTCTTGCTGGTGTGACAGTACAGTCTGGATTAACAAGAC 1500
Qy AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValProLeu 520
Db 1501 AATGCCCATCTTGGTACGACCCCTTTCAGGCTACTGTCGAGAGAGTCCCTTA 1560
Qy GlyTyrLeuValLeuHisValGlnAlaAlaAspAlaAspAlaGlyAspAsnAlaArgLeu 540
Db 1561 GGTCACTGTGTTCTTCAGTGCAGGCTATGACGCTGATGTGTGCAAGTCCCGCTG 1620
Qy GluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThrGly 560
Db 1621 GAATACGCTTGTGCTGGGTGGACATGACTTCCCTTCAACATCAACATGACAGGC 1680
Qy TyrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580
Db 1681 TGGATCTCTGTGGCTGGAATGGAACGGGAGAGATTGATTTCTACAGCTTGGGTA 1740
Qy GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1741 GAAGCTCGAGACCAATGCACTCCAGCACTGCTCGGCAAGTGTCAAGCTGACTGTC 1800
Qy LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620
Db 1801 CTGATGTCAACGACACATCACTTAACCTTTCACCAACAGATGACAGTGGCTCAAT 1860
Qy GluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640
Db 1861 GAGGATCAGCTGTGGGACCAAGGTGTGAAGGTGTGAGTGTGACCGTGAATGCTCAT 1920
Qy SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThrSer 660
Db 1921 AGTATCACTACCTACAGATCAACAGTGGCAATCTGAAACCCCTTCTCATCAACAGC 1980
Qy GlnSerGlyGlyValLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluAlaArgGln 680
Db 1981 CAAGTGTGGTGGCTGGTATCCCTTGCCTGCGCACTGACATCAAACTTGAAGCGGAG 2040
Qy TyrValLeuAlaValAlaThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700
Db 2041 TATGTGTGGCTGTACCGCTCCGATGGCACTCGGACGAGCAAGGACAGATGTGGTG 2100
Qy AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720
Db 2101 AATGTACCCGACGCAACACCCATCGTCTGTTCAGAGCTCCCACTATACATGAT 2160
Qy ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740
Db 2161 GTTATATAGGACCGGCGGGAGGACCAAGTGTGTATCAAGCGCACAGAGAGAGAC 2220
Qy ThrGlyLeuAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760
Db 2221 ACAGGTAGAAATGCCGACATCACTACTTCAATGAGAGACAGCATCCCCAGTTCCGATC 2280
Qy AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780
Db 2281 GATGACAGACCGGGGGCTGTCAACACCAGCTGAGCTGTGATCAAGAACCAAGTGTCT 2340
Qy TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnTyrSerAspThrThrTyr 800
Db 2341 TACACCTGGCACTTACTGTCCGAGACAAAGGCAATCCCAAGATGCCACACCTAC 2400
Qy LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820
Db 2401 CTGGAGATCTGTGGTGAACGATGATGACAAATGCCCTCAGTTCTCGGAGACTCTAC 2460
Qy GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840
Db 2461 CAGGGCAGTGTCTATAGAGATGTCACCCCTTCACTAGCTCTGCAAGATCTCAAGCACT 2520
Qy AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860
Db 2521 GATCGATTTCTGACTTAATGCGAGGGTCTTCACTTCCAAAGAGCGACAGATGA 2580
Qy AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880
Db 2581 GACGGTACTTTATGTGTGAGTCAAGTCAAGCATCGTGGAAACGTACGAGAGCTGAT 2640
Qy ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIleGlyMetProPro 900
Db 2641 CAGAGAACGTGGCCCAATGATGTCTTGGCGCATGACAGTGAACAAGGGAGGCCCA 2700
Qy AlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProProVal 920
Db 2701 GCCCGACACTTATGAAATGACATGCTGCTGTGGATGTGAATGAAATCCCTGTC 2760
Qy PheGlnGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAlaVal 940
Db 2761 TTGACACAGATGATGATGATGTGTGTGAGAGAGAAACGCCCATTTGGGCTAGCGTG 2820
Qy AlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGlnIle 960
Db 2821 GCCCGGGTCAACGCCACTGACCCCGATGAAGGACCAATGCCAGATTAATGATCCAGATT 2880
Qy ValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThrAla 980
Db 2881 GTGAGGGCAATCTCTGAGGTCTTCCAGCTGACATCTTCTCCGGGAGCTGACAGCC 2940
Qy LeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThrSer 1000
Db 2941 CTGGTGAATTAACATACAGAGACCGGCTGAGTACGTCTGTGTATCCAGGCAAGTCA 3000
Qy AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspPro 1020
Db 3001 GCTTCTGTGTGAGCGGGGCTACAGTCAAGTCCGCTCTTCAACGCAATGACCA 3060
Qy ProValLeuGlyAsnPheGluIleLeuPheAsnAspTyrValThrAsnArgSerSer 1040
Db 3061 CAGGTCTGGGCACTTATGATCTTTCACAACTATGTCCATATCCATCCCAACAGC 3120
Qy PheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSerLeu 1060
Db 3121 TTCCCTGGGGGTGCCATTTGGCCGAGTACCTGCCCAAGACCTGATATCTCAGATGCTG 3180
Qy ThrTyrSerPheGluArgGlyValAsnGlyLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080
Db 3181 ACTTACAGCTTTGAGCGGGGAAATGAATCAAGCTGCTGCTCAATGCTTCAAGGAGT 3240

```

QY 1081 GluLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
 DB 3241 GAGCTGAAGCTAAGCGCGCACTGGACAACAACCGGCTCTGGAGGCATCATGAGCGTG 3300  
 QY 1101 LeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1120  
 DB 3301 CTGGGTGACAGCGGGTACACAGCGTGAACCGCCAGTGGCGCTCGGTGTGCATCATC 3360  
 QY 1121 ThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGluArg 1140  
 DB 3361 ACCGATGAGATCTCAACCAAGACATCAAGCTGGCCCTGGAGAACATGTCAACCGAGGCG 3420  
 QY 1141 PheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThrPro 1160  
 DB 3421 TTCCTGTACCACTGCTAGGCTCTTCATCCAGGGGTGGCCGCCACCGTGGCCACGCGCA 3480  
 QY 1161 ProAspHisValValValPheAsnValGlnArgAspThrAspAlaProGlyGlyHisIle 1180  
 DB 3481 CCGAACCAAGTGTGTCTTCAACGATACAGCGGGAACCGACGCCCGGGGGCCACATC 3540  
 QY 1181 LeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProProPheLeu 1200  
 DB 3541 CTCAACGTAGGCTGTCTGGTGGCCAGCCGACGAGGCCCGGGGGGCGCGCTTCTCG 3600  
 QY 1201 ProSerGluAspLeuGlnGluArgLeuTyLeuAsnArgSerLeuLeuThrAlaIleSer 1220  
 DB 3601 CCTCTGAAGCACTGCAAGAGCGCTTACCTCAACCGGACCTGCTGACGCGCATCTCG 3660  
 QY 1221 AlaGlnArgValLeuProPheAspAsnIleCysLeuArgGlnProCysGluAsnTyr 1240  
 DB 3661 GCAACGCGCGTGTCTCTTCAACCAACATCTCTGCGGAGCCCTGCGACAGATCAC 3720  
 QY 1241 MetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSerSer 1260  
 DB 3721 ATGGCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 3780  
 QY 1261 ValLeuPheArgProIleHisArgProValGlyLeuArgCysArgCysProProGlyPhe 1280  
 DB 3781 GTGCTCTTCCGGCCATCAACCCCGGAGGGGTGCGCTGCGCTGCGCCCGCGCTTC 3840  
 QY 1281 ThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyProHis 1300  
 DB 3841 ACGGTGTACTGTGAGACCGAGGTGTGACCTGTCTACTGCGCGCCCTGTGTGGCCCCAC 3900  
 QY 1301 GlyArgCysArgSerArgGluGlyGlyTyrThrCysLeuCysArgAspGlyTyrThrGly 1320  
 DB 3901 GGGCGCTGCGGACCGGAGGGCGGCTTACCTGCTGTGTGTGTGTGTGTGTGTGTGTGT 3960  
 QY 1321 GlnHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsnGly 1340  
 DB 3961 GAGCACTGTGAGGTGAGTCTGCTCAGGCGGTGTCAACCGGAGGTGTGTGTGTGTGTGT 4020  
 QY 1341 GlyThrCysValAsnLeuLeuValGlyGlyPheGlyCysAspCysProSerGlyAspPhe 1360  
 DB 4021 GGCACCTGTGTAACTGT 4080  
 QY 1361 GlyLysProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThrPhe 1380  
 DB 4081 GAGAGGCGCTTATGCTCAAGGTGTCAACCGGCACTTCCCGGCCCATCTCTTCACTCACTT 4140  
 QY 1381 ArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGluArg 1400  
 DB 4141 CCGCGCTGCGCACGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 4200  
 QY 1401 AspGlyLeuLeuLeuThrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeuGlu 1420  
 DB 4201 GAGGGGT 4260  
 QY 1421 ValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrValSer 1440  
 DB 4261 GTGATCCAGGAGCAAGGTCAAGCTTCTTGTGAGGGAGTCAACCAACCAACGAGTGTCTC 4320

QY 1441 ProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGlnLeuLysTyrTyr 1460  
 DB 4321 CCAATTCGTGCGCGGAGAGATCATGTGATGGCCAGTGGCATACGCTGTGAATATCTTAC 4380  
 QY 1461 AsnLysProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnValAla 1480  
 DB 4381 AATAAGCCACTGT 4440  
 QY 1481 ValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGly 1500  
 DB 4441 GTGGTGAACGGTGAAGT 4500  
 QY 1501 AsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerLysLysSerLeuAspLeuThr 1520  
 DB 4501 AACTACTCTGT 4560  
 QY 1521 GlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMetArg 1540  
 DB 4561 GGGCCCTGTCTACAGGCGGGGT 4620  
 QY 1541 GlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAsp 1560  
 DB 4621 CAGTTCGTGGCTGTCAATGCGGAACCTCAAGTGTGACCGGCAATAGATGTGCTGAC 4680  
 QY 1561 PheIleAlaAsnAsnGlyThrValProGlyCysProAlaLysLysAsnValCysAspSer 1580  
 DB 4681 TTCAATGCCAACAATGTGCACCGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4740  
 QY 1581 AsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGluCys 1600  
 DB 4741 AACCTTGCCACAATGTGGGGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
 QY 1601 ProLeuGlyPheGlyGlyLysSerCysAlaGlnGlnMetAlaAsnProGlnHisPheLeu 1620  
 DB 4801 CCTTGGGCTTGTGGGGGCAAGCTGTGGCCCAAGAAATGTGCATTCACAGCACTTCTCG 4860  
 QY 1621 GlySerSerLeuValAlaThrPheGlyLeuSerLeuProIleSerGlnProTyrTyrLeu 1640  
 DB 4861 GGCAGCAGCTGT 4920  
 QY 1641 SerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGly 1660  
 DB 4921 AGCTCAATGTTCGACAGCGCCGACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980  
 QY 1661 ArgSerThrIleThrLeuGlnLeuArgGluGlyHisValMetLeuSerValGluGlyThr 1680  
 DB 4981 CGCAGCAACATCAACCTTCAAGCTTCAAGAGGGCCACTGATGTCAAGCTGTGAGGGACA 5040  
 QY 1681 GlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyValArgAlaAsnAspGlyAspThrHis 1700  
 DB 5041 GGGCTTCAAGGCTTCTCTCTCCGTGTGAGCCAGGCGGGCCCAATGTACGCTGTGCAC 5100  
 QY 1701 HisAlaGlnLeuAlaLeuGlyAlaSerGlyGlyProGlyHisAlaIleLeuSerPheAsp 1720  
 DB 5101 CATGACAGCTGTGAGCTGTGAGCGGAGCGGAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 5160  
 QY 1721 TyrGlyGlnGlnArgAlaGluGlyAsnLeuGlyProArgLeuHisGlyLeuHisLeuSer 1740  
 DB 5161 TATGGGACGAGAGAGAGAGAGAGGAGCACTGGGCCCGGGCTGTGATGTGTGTGTGTGTGT 5220  
 QY 1741 AsnIleThrValGlyGlyIleProGlyProAlaGlyGlyValAlaAspGlyPheArgGly 1760  
 DB 5221 AACATTAACAGTGGCGGAAATACCTGTGGCCAGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 5280  
 QY 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
 DB 5281 TGTTTGAGGGGT 5340  
 QY 1781 HisGlyLysSerIleAsnValGlnGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
 DB 5341 CATGGGAGAGCATCAACGTGTGAGCAAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400  
 QY 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyrSerCysSerCysAsp 1820

Db 5401 CCGGTCCTGCTAACAGCTATTTGACAGACGAGCTGGACAGCTATTCCTGACGCTGAT 5460  
 Qy 1821 ProGlyTyrTyrGlyAspAsnCySerThrAsnValCysAspLeuAanProCysGluNH<sub>2</sub>Gln 1840  
 Db 5461 CAGGTTACTAGTGTGACACTGACTACTATGTGTGTACTGAAACCCGTGTGACACAG 5520  
 Qy 1841 SerValCysThrArgIysProSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
 Db 5521 TCTGTGTGTATCCCGCAAGCCAGGCGCCCATGAGTATACCTGCAAGTGTCCCAAT 5580  
 Qy 1861 TyrLeuGlyProTyrCysGluThrArgGlyAspGlnProCysProArgGlyTyrTyrGly 1880  
 Db 5581 TACCTTGAGGCATACCTGTGAGACCAAGATTCAGACGCTTGCCCTGTGCTGTGGGA 5640  
 Qy 1881 HisProThrCysGlyProCysAsnCyAspValSerIysGlyPheAspProAspCysAsn 1900  
 Db 5641 CATCCCAATGTGGCCCATGACATGATGATGTCGACAAAGGCTTTGACCCAGACTGCAC 5700  
 Qy 1901 LysThrSerGlyGluCysHisCysValGluAsnHisTyrArgProProGlySerProThr 1920  
 Db 5701 AAGCAAGCCGCGAGTGCCTGCAAGAGAACCACTACCGGCCCCAGGCAAGCCCCACC 5760  
 Qy 1921 CysLeuLeuCyAspCysArgTyrProThrGlySerLeuSerArgValCysAspProGlyAsp 1940  
 Db 5761 TGCCTCTGTGTGACCTGACCTCCAGGCTCTTGTCCAGAGCTGTGTACCTTGAGAT 5820  
 Qy 1941 GlyGlnCysProCysIysProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
 Db 5821 GGCACAGTGTCCAGCAAGCCAGGTGTATGGGGGTGATGATGACCGCTGTGACACCT 5880  
 Qy 1961 PheAlaGluValThrThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle 1980  
 Db 5881 TTGTGTGAGGTCCACCAATGGCTGTGAAGTGAATGAACTGACAGCTGCCACAGCGCAT 5940  
 Qy 1981 GluAlaGlyIleTyrTyrProArgTyrArgPheGlyLeuProAlaAlaAlaProCysPro 2000  
 Db 5941 GAGGCTGGAGATCGAGGCGCCGACCGGCTGCGGCTGTGCTGTCTCTGCTGCC 6000  
 Qy 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGlyLysArgGlyTyrLeuProPro 2020  
 Db 6001 AAGGCTCTTTGGAGCTGCTGTGGCCCATGTGTATGACAGGGGGTGGCTCCCCCA 6060  
 Qy 2021 AsnLeuPheAsnCySerThrSerIleThrPheSerGlyLeuLysGlyPheAlaGluArgLeu 2040  
 Db 6061 AACCTCTTCAACTGACGTCATCACCTTCTCAGAACTGAAGGGCTTCGTGAGCGGCTA 6120  
 Qy 2041 GluArgAsnGlySerGlyLeuAspSerGlyArgSerGlnIleuAlaLeuLeuLeuArg 2060  
 Db 6121 CAGCGAAATGATGACAGGCTTAGACTCAGGGGCGCTCCAGCAGCTAGCCCTGCTGCC 6180  
 Qy 2061 AsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIysValAlaTyrGlnLeu 2080  
 Db 6181 AAGCCACGAGACACAGCTGCTACTTCTGGCAGGCACTCAAGGTGGCTTACACAGCTG 6240  
 Qy 2081 AlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGln 2100  
 Db 6241 GCCACCGCGGTGTGGCCCAAGAGACCCACAGCGGGCTTGTGGGCTGTGCCACACAG 6300  
 Qy 2101 AspValHisPheThrGlnLeuLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
 Db 6301 GAGGTGACTTCACTGAAATGTGTGGGTGGGAGCGCCCTCTGACACAGCAAC 6360  
 Qy 2121 LysArgHisTyrPheLeuLeuIleGlnGlnThrGluGlyValThrAlaTyrLeuLeuGlnHis 2140  
 Db 6361 AAGCGGCACTGGAGCTGATCCAGCAGACAGAGGGTGGACCGCTGTGGCTGCCAGCAGC 6420  
 Qy 2141 TyrGluAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe 2160  
 Db 6421 TAGGAGCCCTAACCCAGTGTCCGTGGCCCAAGACATGTGGGCAACCTAACGCCCTTTC 6480  
 Qy 2161 ThrIleValThrProAsnIleValIleSerValIleArgLeuAspLysGlyAsnPheAla 2180

Db 6481 ACCATGTACAGCCCAACATTTGTACCTCCGTAGTGTGCTTGGACAAAGGAATTTGCT 6540  
 Qy 2181 GlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGlnGlnProProAspLeuGluThr 2200  
 Db 6541 GGGGCCAGCTTGGCCCGCTTACGAGGCCCTGTGGGGAGAGAGCCCCGGGACTTGGACA 6600  
 Qy 2201 ThrValIleLeuProGlySerValPheArgGluThrProProValValArgProAlaGly 2220  
 Db 6601 ACAGTCAATTCCTGAGTGTGTCTTCAAGAGAGCCCGCCGTGTGACGGCCGACAGC 6660  
 Qy 2221 ProGlyGluAlaGlnGluProGlyGluLeuAlaArgArgGlnArgArgHisProGlyLeu 2240  
 Db 6661 CCCGAGAGGCCAGAGAGCCAGAGGAGCTGGCACCGCCGACAGGAGAGGCCGAGCTG 6720  
 Qy 2241 SerGlnGlyLysAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 6721 AGCCAGGGTGAAGCTTGGCCAGCTCATCATCTTACCGACCTGTGGCGGCTACGTGCT 6780  
 Qy 2261 HisAsnTyrAspProAspLysArgSerLeuArgValProLysArgProIleIleAsnThr 2280  
 Db 6781 CATPACTATGACCTTGACAGAGGCACTGTGAGGTCCCAAGGCCGATCATCACACA 6840  
 Qy 2281 ProValIleSerIleSerValHisAspArgGlyGluLeuLeuProArgAlaLeuAspLys 2300  
 Db 6841 CCGGTGTGAGCATCAGCTTCATGATGATGAGAGCTTCTGCCCCGGGCCCCGACAA 6900  
 Qy 2301 ProValThrValGlnPheArgLeuGluThrGlnGluArgTyrThrProIleCysVal 2320  
 Db 6901 CCGGTACGGTGAAGTTCGGCTTGTGAGACAGAGAGCCGAGCCAGCTGTGTGTC 6960  
 Qy 2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGlu 2340  
 Db 6961 TTCTGTGAACATTCATCTGTGTGAGTGGCACAGGTGGCTGTGGCCGACAGAGCTGTGA 7020  
 Qy 2341 ValValPheArgAsnGlySerHisValSerCysGlnCysAsnHisMetThrSerPheAla 2360  
 Db 7021 GTGTCTTCCGAAATGAGAGCCAGCTCAGCTTCAGTGAACCACTGACAGCTTCGCT 7080  
 Qy 2361 ValLeuMetAspValSerArgArgGluAsnGlyGluIleLeuProLeuYsrThrLeuThr 2380  
 Db 7081 GTGTCTATGAGCTTTCTTCGGCGGAGAGATGGGAGATCTGCCACTGAAACACTGACA 7140  
 Qy 2381 TyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuThrPhePhePheLeuThrLeu 2400  
 Db 7141 TAGTGGCTCTAGGTGTCACCTTGGCTGGCTTGTGCTCACTTCTTCTCTCACTTC 7200  
 Qy 2401 LeuArgGlyLeuArgSerAsnGlnHisGlyIleArgArgAsnLeuThrAlaAlaLeuGly 2420  
 Db 7201 TGGCGATCTGTGCTCCAAACACAGGATCCGACCTTACCTTGAACAGCTGCCCTGGGC 7260  
 Qy 2421 LeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThr 2440  
 Db 7261 CTGGCTCAGCTGTCTTCTCTCGGGATCAACAGGCTGACCTCCCTTTTGGCTGGACA 7320  
 Qy 2441 ValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTyrAlaLeuLeuGln 2460  
 Db 7321 GTCATTTGCCATCTGTGACTTCTCTTACCTCTGACACTTCTTCTGGGCTGTGCGAG 7380  
 Qy 2461 AlaLeuHisLeuTyrArgAlaLeuThrGluValArgAspValAsnThrGlyProMetLeu 2480  
 Db 7381 GCTTTGCACTGTATACCGGGCACTCACTAGGTGCGAGATGACACCGGCCCAATGGCC 7440  
 Qy 2481 PheTyrTyrMetLeuGlyTyrProGlyValProAlaPheIleThrGlyLeuAlaValGlyLeu 2500  
 Db 7441 TTCTACTACATGCTGGGCTGGGCGCTGTGCTGTCTTCAACAGGCTAGCCGTGGGCTG 7500  
 Qy 2501 AspProGluGlyTyrGlyAsnProAspPheCysTyrPheSerIleTyrAspThrLeuIle 2520  
 Db 7501 GACCCCGAGGGGTACGGGAACTGTACTGTGCTGTGCTCTCATCTATGACAGCTCATC 7560  
 Qy 2521 TrpSerPheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyrIleLeu 2540  
 Db 7561 TGGAGTTTGTGCGCCCGGTGCTTGTGCTGTGATGAGTGTCTTCTGTACATCTGTG 7620

QY 2541 AAlaAaArgAlaSerCyAlaAlaGlnArgGlnGlyPheGlyLysLysGlyProValSer 2560  
 Db 7621 GCGGCCCGGGGCTCTGTCGTGCTCCAGCGGCGAGGGCTTGAAGAAAGAGTCTGTCTCG 7680  
 QY 2551 G1LysGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeuAlaLeu 2580  
 Db 7681 GGGCTGCGAGCCCTCTGCGCGTCTGCTGCTGAGCGCCAGTGTGCTGTGGCACTG 7740  
 QY 2581 LeuSerValaAsnSerAspThrLeuLeuPheHisTrpLeuPheAlaThrCyAsnCyAla 2600  
 Db 7741 CTCTGTGCAACAGGACACCTCTCTTCCACTACCTTGTGCTGACCTGCAATTGCAATC 7800  
 QY 2601 G1nGlyProPheAlaPheLeuSerTrpValValLeuSerLysGlyValArgLysAlaLeu 2620  
 Db 7801 CAGGGCCCTTATCTTCTCTCTCTATGTGTGCTTACCAAGAGTCCGGAAGACATC 7860  
 QY 2621 LysLeuAlaCySerArgLysProSerProAspProAlaLeuThrThrLysSerThrLeu 2640  
 Db 7861 AAGCTTGCTGCGAGCGGAGCGCCAGCCCTGACCTGCTGAGCAACCAAGTCCACCTG 7920  
 QY 2641 ThrSerSerTrpAsnCyProSerProTrpValaAspGlyArgLeuTrpGlnProTrpGly 2660  
 Db 7921 ACCTGCTCTACACTGCCCCAGCCCTTACGCAAGTGGCGGCTGTACAGCCCTACCGGA 7980  
 QY 2661 AspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSerGlnProSerTrpLe 2680  
 Db 7981 GACTGGCGCGGCTCTGTCACAGCACACAGTGTGCTGGGCAAGAGTCAACCCAGCTACATC 8040  
 QY 2681 ProPheLeuLeuArgGlnGlnSerAlaLeuAsnProGlyGlnGlyProProGlyLeuGly 2700  
 Db 8041 CCTCTCTGTGAGGAGGAGAGTCCGCACTGAACTGTGCGCAAGGGCCCCCTGCGCTGGGG 8100  
 QY 2701 AspProGlySerLeuPheLeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSer 2720  
 Db 8101 GATTCAGGACAGCTGTCTCTGAGAGTCAAGACAGCAAGTATCTGACACGGACTCC 8160  
 QY 2721 AspSerAspLeuSerLeuGlnAspAspGlnSerGlySerTrpAlaSerThrHisSerSer 2740  
 Db 8161 GACAGTGAAGCTGTCTTGAAGAGACAGAGTGTGCTCTATGCTCTACCACTCATCA 8220  
 QY 2741 AspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2760  
 Db 8221 GACAGTGAAG 8280  
 QY 2761 TrpAspSerLeuLeuGlyProGlyValaGlyLeuArgLeuProLeuHisSerThrProLysAsp 2780  
 Db 8281 TGGAGTACCTGTCTGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 8340  
 QY 2781 G1yGlyProGlyProGlyLysAlaProTrpProGlyAspPheGlyThrThrAlaLysGlu 2800  
 Db 8341 GGGGGCCCGAGGCTGGGAGAGGCCCCCTGGCGAGAGACTTTGGGACCAAGCAAAAGAG 8400  
 QY 2801 SerSerGlyAsnGlyValaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2820  
 Db 8401 AGTATGTGGAG 8460  
 QY 2821 G1nGlySerLeuGlyProLeuProGlySerSerAlaGlnProHisLys----- 2836  
 Db 8461 GAGGGGTCTTGAAGCCCCCTTCCAGGCTTCTGCCCCAGCTTCAAAAGGTGAGTGGGGC 8520  
 QY 2836 ----- 2836  
 Db 8521 ACCCCAGCTGCGAGAGCTCCCTAGTACAGAGCTCATATCTCATATTCCTGTGGCG 8580  
 QY 2837 -----G1LysLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 2847  
 Db 8581 CACTCAGAGCCCG 8640  
 QY 2848 SerGlyLysSerSerLeuLeuArgLeuProLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 2867  
 Db 8641 AGCGAG 8700

QY 2868 SerSerAlaSerGlnGlySerArgGlyGlyProProProArgProProProArgGlnSer 2887  
 Db 8701 TCTTCGCTAGTGAAGGCGAGCCCGGGCGGCCCCCTCCCGCCACCGCCCCGCGAGAGC 8760  
 QY 2888 LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2907  
 Db 8761 CTCAG 8820  
 QY 2908 AspGlnAspSerSerGlySerGlnPheLeuPhePheAsnPheLeuHis 2923  
 Db 8821 GATGAGACTGTGAGGCTCGAATTTCTTCTTAATCTCTGAT 8868  
 RESULT 4  
 US-10-292-798-931  
 : Sequence 931, Application us/10292798  
 : Publication No. US20030235833A1  
 : GENERAL INFORMATION:  
 : APPLICANT: SUMA, MAKIKO  
 : APPLICANT: ASAI, KIYOSHI  
 : APPLICANT: AKIYAMA, YUTAKA  
 : APPLICANT: ABURATANI, HIROYUKI  
 : TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
 : FILE REFERENCE: 084335/166  
 : CURRENT FILING DATE: 2002-11-13  
 : PRIOR APPLICATION NUMBER: 10/017,161  
 : PRIOR FILING DATE: 2001-12-18  
 : PRIOR APPLICATION NUMBER: JP 2001-246789  
 : NUMBER OF SEQ ID NOS: 2070  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 931  
 : LENGTH: 24370  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : LOCATION: source  
 : FEATURE: (1)..(24370)  
 : NAME/KEY: CDS  
 : LOCATION: (201)..(3510)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (8553)..(9200)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (11163)..(11385)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (11634)..(11749)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (11929)..(12017)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (12408)..(12565)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (12927)..(13090)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (13263)..(13388)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (13732)..(13898)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (14199)..(14486)  
 : FEATURE:  
 : NAME/KEY: CDS  
 : LOCATION: (14575)..(14749)  
 : FEATURE:

NAME/KEY: CDS  
 LOCATION: (14989) .. (15130)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (15312) .. (15432)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (15856) .. (16039)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (16226) .. (16327)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (17669) .. (17763)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (17972) .. (18181)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (18702) .. (18885)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19001) .. (19129)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19231) .. (19411)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19545) .. (19712)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (19815) .. (19941)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (20054) .. (20254)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (20547) .. (20721)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21048) .. (21160)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21338) .. (21417)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21507) .. (21633)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (21721) .. (21843)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22399) .. (22526)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22761) .. (22849)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (22954) .. (23148)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (23287) .. (23457)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (23557) .. (23791)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (24143) .. (24170)  
 US-10-292-798-931

Alignment Scores:  
 Pred. No.: 0  
 Score: 12339.00

Length: 24370  
 Matches: 2907

Percent Similarity: 36.92%  
 Best Local Similarity: 36.92%  
 Query Match: 79.38%  
 DB: 12  
 Gaps: 32  
 US-09-916-849a-3 (1-2923) x US-10-292-798-931 (1-24370)

QY 1 MetArgSerProAlaThrGlyValProLeuProThrProProProLeuLeuLeu 20  
 DB 201 ATGGAGAGCCGGCCACCGGGCTCCCTCCCAAGCCCGCGCTGCTGCTGCTG 260

QY 21 LeuLeuLeuLeuLeuProProProLeuLeuGlyAspGlnValGlyProCysArgSerLeu 40  
 DB 261 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320

QY 41 GlySerArgGlyArgGlySerSerGlyValAlaProMetGlyThrLeuCysProSer 60  
 DB 321 GGATCCAGGAG 380

QY 61 SerAlaSerAsnLeuTrpLeuTrpThrSerArgCysArgAspAlaGlyThrGlnLeuThr 80  
 DB 381 TCAGGCTCGAACCTTGCTGCTTACACAGCCGCTGAGGAGATCGGCACTGAGCTGAT 440

QY 81 GlyHisLeuValProHisHisAspGlyLeuAlaGlyValTrpCysProGlySerGlyAlaHis 100  
 DB 441 GGCCACCTGTCACCCACACAGATGCTGAGGGTTGGTGTTCAGAAATCCAGAGCCCAT 500

QY 101 IleProLeuProProAlaProGlyGlyCysProTrpSerCysArgLeuLeuGlyIleGly 120  
 DB 501 ATTCCTCTTCACACAGCTCTTGAAGCTCCCTGAGCTGTGCTCTGCGCATTTGA 560

QY 121 GlyHisLeuSerProGlnGlyLysLeuThrLeuProGlnGlyHisProCysLeuGlyAla 140  
 DB 561 GGCCACCTTTCCCAAGAGGCAAGCTCACATGCCCCGAGAGAGACCCGATTAAGGCT 620

QY 141 ProArgLeuArgCysGlnSerCysLysLeuAlaGlnAlaProGlyLeuArgAlaGlyGlu 160  
 DB 621 CCACGGCTCAGATGCAATGCTGCAAGCTGCAAGGCCCCCGGCTCAGGGAGGGAGA 680

QY 161 ArgSerProGlnGlySerLeuGlyGlyArgGlyLysArgAsnValAsnThrAlaProGln 180  
 DB 681 AGCTACACAGAAAGTCCCTGGGTGGGCTCCGAAAGAAATGTAAATCAGCCCCCAG 740

QY 181 PheGlnProProSerTyrglnAlaThrValProGlnAsnGlnProAlaGlyThrProVal 200  
 DB 741 TTCCAGCCCCCAGCTACCAAGGCCACAGTCCGAGAACCAAGCAGCAGCCCTGTT 800

QY 201 AlaSerLeuArgAlaIleAspProAspGlyGlyAlaGlyArgLeuGlyTrpThrMet 220  
 DB 801 GCATCCCTGAGGACCATTCAGCCGAGCAGAGGTGAGGAGTGCAGCTGAGTACACCATG 860

QY 221 AspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProValThrGlyAla 240  
 DB 861 GATGCCCTCTTTGATAGCCGCTCCACACAGTTCTTCTCCCTGAGCCCATCTGATGA 920

QY 241 ValThrThrAlaGlnGlnLeuAspArgGlyTrpThrLysSerThrHisValPheArgValThr 260  
 DB 921 GTAAACCAAGCCGAGAGAGCTGATGCTGAGACCAAGAGCAACCAAGCTTCAAGGGTCA 980

QY 261 AlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThr 280  
 DB 981 GGCAAGAGACAGGAG 1040

QY 281 AspThrAsnAspHisAspProValPheGlnGlnGlnGlyTrpLysGlnSerLeuArgGlu 300  
 DB 1041 GACACCAATGACCAATGACCTGTGTTCCAGACAGAGAGTACAAGAGAGAGAGAGAG 1100

QY 301 AsnLeuGlnValGlyTyrglnValIleuThrValAlaGlyAlaThrAspGlyAspAlaProPro 320  
 DB 1101 AACCTGAGGTGGCTATGAGGTGCTCACTGTCAGGGCCACGAGTGTATATCCCTCC 1160

QY 321 AsnAlaAsnIleLeuTrpArgLeuLeuGlnGlySerProSerGlyValPhe 340

Db 1161 AATGCCAATATCTGTACCGCTGCTGAGGGGCTCTGGGGGACGCCCTCTGAAGTCTTT 1220  
Qy 341 GIUIIAspProArgSerGIYValIleArgThrArgGIYProValAspArgGIUVal 360  
Db 1221 GAGATCGACCTCGCTCTGGGGTGATCCGAACCCGCGGCTCTGTGATCGGGAAGAGGTG 1280  
Qy 361 GIUSerTYrGlnLeuThrValGIUValAspArgGlnGIYArgAspProGIYProArgSer 380  
Db 1281 GAATCCATCAAGCTGACGGGTAGAGGCAGGTGACAGGGGTGGGACCCGGGTCTCGAGT 1340  
Qy 381 ThrThrAlaIleValPheLeuSerValGIUAspAspAspAspAlaProGlnPheSer 400  
Db 1341 ACCACAGCCGCTCTTTCTTTCTCTGTGAGAGTACATGATGATAGTCCCGAGTTTACT 1400  
Qy 401 GIUArgTYrValValGIUValArgGIUAspValThrProGIYAlaProValLeuArg 420  
Db 1401 GAGAAAGCGCTATGTTGTCAGAGGTGAGGAGATGTGATCCAGGGGCCCAAGTATCCGA 1460  
Qy 421 ValThrAlaSerAspArgAspIYSGIYSerAspAlaValIleIleTYrSerIleMetSer 440  
Db 1461 GTACAGCCTCGGATCGAGACAAAGGAGCAATGCCGTGTGCTACTATAGCATGTAGT 1520  
Qy 441 GIYAspAlaArgGIYGIUAspValArgIleuAspAlaGlnThrGIYAlaLeuAspValIleSer 460  
Db 1521 GGCATATGCTCGGGGACAGTTTATCTGATGCCAGACTGAGAGCTCTGGATGTGTGAGC 1580  
Qy 461 ProLeuAspTYrGIUThrThrIYSGIYTYrIleuArgValArgAlaGIUAspGIYGIY 480  
Db 1581 CCTCTTGACTATGAGACGACCAAGAGATGACCTTACGGGTGCGAGCACAGAGTGTGTGC 1640  
Qy 481 ArgProProLeuSerAspValSerGIYLeuValThrValGlnValLeuAspIleAspAsp 500  
Db 1641 GGTCCCACTCTTAATGCTCTGCTGTGTGAGACATGACAGTCTGTGATATCAACAC 1700  
Qy 501 AsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGIUAspValProLeu 520  
Db 1701 AATGCCCATCTTGTCAGACACCCCTTCCAGGCTACTGCTCGAGAGTGTCCCTTA 1760  
Qy 521 GIYTYrLeuValLeuHisValGIUValIleAspAlaAspAlaGIYAspAspAlaArgLeu 540  
Db 1761 GGTCTACCTGTCTCTCATGTCCAGGCTATCGACGTGTGTGTGAGACATGCCGCTG 1820  
Qy 541 GIUTYrArgLeuAlaGIYValGIYHisAspPheProPheThrIleAspAspGIYTYrGIY 560  
Db 1821 GAATACCGCTTGCTGGGGTGAGACATGCTTCCCTTCCATCAACATGACATGACAGC 1880  
Qy 561 TrpIleSerValAlaIleGIUAspArgGIUValIleAspPheTYrSerPheGIYVal 580  
Db 1881 TGGATCTGTGTGCTGTGAATGGAACCGGAGAGATTGATTTCTACAGCTTTGGGATA 1940  
Qy 581 GIUAlaArgAspHisGIYTYrProAlaLeuThrAlaSerAlaSerValIleThrVal 600  
Db 1941 GAAGCTCGAGCAATGCGACTCCAGACACTCATGCTCGGCGAGGTGACGCTGACTGTC 2000  
Qy 601 LeuAspValAspAspAspAspProThrPheThrGlnIleProGIUValTYrThrValArgLeuAsn 620  
Db 2001 CTGATGTCAACAGCAACATCAACCTTACCAACAGAGTACACAGTGGGCTCAT 2060  
Qy 621 GIUAspAlaIleValGIYTYrSerValIleThrValIleSerAlaValAspArgAspAlaHis 640  
Db 2061 GAGAGTGAAGCTGTGGGACCAAGCGTGTGACGCTGTGACCGTGTGATCTCAT 2120  
Qy 641 SerValIleThrTYrGlnIleThrSerGIYAsnThrArgAspAspPheSerIleThrSer 660  
Db 2121 AGTGTCACTACCTACAGATCAACAGTGGCAATATCTCGAACCCTTCTCCATCAACGAC 2180  
Qy 661 GlnSerGIYGIYGIYLeuValSerLeuAlaLeuProLeuAspTYrTYrLeuGlnArgGln 680  
Db 2181 CAAGGTGTGTGTGTGTATCTTGTCTGTGCACTGGACTTACAACTTGAGGGGACG 2240  
Qy 681 TYrValIleuAlaValThrAlaSerAspGIYTYrArgGlnAspThrAlaGlnIleValVal 700  
Db 2241 TATGTGTGTGTGTGTATCCGCTCGAGTGGCACTCGGAGGACACAGGACAGATTTGTGTG 2300

Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTYrThrValAsn 720  
Db 2301 AATGTACCAAGCGCCAAACCAACATGCTCTGTCTTTCAAGAGCTCCCATATACGTGAT 2360  
Qy 721 ValAsnGIUAspArgProAlaGIYTYrThrValValLeuIleSerAlaThrAspGIUAsp 740  
Db 2361 GTTATATAGGACCGGCGGAGGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2420  
Qy 741 ThrGIYGIUAspAlaArgIleThrTYrPheMetGIUAspSerIleProGlnPheArgIle 760  
Db 2421 ACAGTGTAGATGCCCGCACTACCTACTTCAATGAGACAGATCCCGCAAGTGTGTCT 2480  
Qy 761 AspAlaAspThrGIYAlaValThrThrGlnAlaGIUAspAspTYrGIUAspGlnValSer 780  
Db 2481 GATGACAGACAGGGGGCTGTACCAACCAAGCTGAGCTGTGATGTGAAGACCAAGTGTCT 2540  
Qy 781 TYrThrIleuAlaIleThrAlaArgAspAspGIYIleProGlnIYSGIYSerAspThrThrTYr 800  
Db 2541 TACACCTGTGCTATGCTGTGGGACATGTGCAATGCCATTCGAGAGTCCGACACCTTAC 2600  
Qy 801 LeuGIUValLeuValAspAspValAspAspAspAlaProGlnPheLeuArgAspSerTYr 820  
Db 2601 CTGAGATTCCTGTGTGAGACAGCTGATGACATGCAATGCCCTCAAGTCTCTGCAAGTCTTAC 2660  
Qy 821 GlnGIYSerValTYrGIUAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2661 CAGGGCAGTGTCTATGAGATGTGTGCCACCTTCACTGACCTGTGAGATCTGAGCCACT 2720  
Qy 841 AspArgAspSerGIYLeuAspGIYArgValPheTYrThrPheGlnGIYGIYAspAspGIY 860  
Db 2721 GATGTATTTCTGTGATTAATGTGAGGGGTCTTCAACCTTCCAGAGAGCGACAGATGGA 2780  
Qy 861 AspGIYAspPheIleValGIUAspThrSerGIYIleValArgThrLeuArgArgLeuAsp 880  
Db 2781 GAGGTGATCTTATTTGTTGATGTGACAGCTCAAGGATGCTGTGAGACGATCGGAGCTGTGAT 2840  
Qy 881 ArgGIUAsnValAlaGlnTYrValLeuArgAlaTYrAlaValAspIYSGIYMetProPro 900  
Db 2841 CGAGAGAACGTGTGCCGTATGTCTTGTGGGCAATGCAATGTGACAAAGGGAGATGCCCA 2900  
Qy 901 AlaArgThrProMetGIUValThrValThrValLeuAspValAspAspAspProProVal 920  
Db 2901 GCCCGCACTTATGAGATGACATGACATGCTGTGTGTGATGTGATGATGATGATGATGAT 2960  
Qy 921 PheGIUValAspGIUAspPheAspValPheValGIUValAsnSerProIleGIYLeuAlaVal 940  
Db 2961 TTTGAGCAGATGTGATGT 3020  
Qy 941 AlaArgValThrAlaThrAspProAspGIYTYrAspAlaGlnIleMetTYrGlnIle 960  
Db 3021 GCCCGGTGACAGGCACTGACCCCGATGAAGGACCAATGCCAGATTATGTATGACATTT 3080  
Qy 961 ValGIUValAsnIleProGIUValPheGlnLeuAspIlePheSerGIYGIULeuThrAla 980  
Db 3081 GTGAGGGCAACATCTCTGAGAGTCTTACAGTGTGATCTTCTCCGGGAGCTGACAGCC 3140  
Qy 981 LeuValAspLeuAspTYrGIUAspArgProGIUValTYrValLeuValIleGlnAlaThrSer 1000  
Db 3141 CTGTGATCACTTATGACTACGAGACCGGCTGAGTACGCTGTGATCCAGGCAAGCTCA 3200  
Qy 1001 AlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspPro 1020  
Db 3201 GCTCTCTGTGTGAGCCGGGTGTACAGTCAAGTGTGCTCTCTTGAACCGCAATGACAAACCA 3260  
Qy 1021 ProValLeuGIYAspPheGIUValLeuPheAspAspTYrValThrAsnArgSerSer 1040  
Db 3261 CCAAGTGTGGCAACTTGTGATCTTTTCAACATATGTACCAATTCCTCAAGACG 3320  
Qy 1041 PheProGIYGIYAlaIleGIYArgValProAlaHisAspProAspIleSerAspSerLeu 1060  
Db 3321 TTCCTGTGGGGGTGCAATGTGACCGAGTACCTGCGCATGACCTGATATCTCAGATAGTCTG 3380

QY 1061 ThrTyrSerPheGluArgIlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThrGly 1080  
Db 3381 ACTTACAGCTTTGAGCGGGAAATGAATCAAGCTGGTCCGTCTCATATGCTTCCACGGGT 3440  
QY 1081 GluLeuIlySerLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSerVal 1100  
Db 3441 GAGCTGAGCTTAAACCCCGCACTGGACAAACACCGGCTCTGGAGGCCATCATGACGGTG 3500  
QY 1101 LeuValSer----- 1103  
Db 3501 CTGGTGTTC-AGGTAAAGAAAGGCCCAAGGTGGCGCTGGGGGTGGGGTAACTCGCGGGATG 3559  
QY 1103 ----- 1103  
Db 3560 GTCTGGGCAACCACTGAGAGTAGGGTGGATCCAGAAAGCACTACAGATCCACTCCCT 3619  
QY 1103 ----- 1103  
Db 3620 GCCCAGTGCTTGACACAGAGCAGAGGGGCAAGAGCCAGCTTGGGAAAGAACCCAGAAA 3679  
QY 1103 ----- 1103  
Db 3680 CCTTGGGGGCTGGTGGCAGAGCCCCACCTTCTGCTGGGCTGGGCTGGCTTCCGTTGGC 3739  
QY 1103 ----- 1103  
Db 3740 TGCCCCCACTGTCTACAGACCTGTGACGTGGTGGGGAAAGTGTGTGAGGCTGCCCGGCG 3799  
QY 1103 ----- 1103  
Db 3800 TATGGCTGCCAAGATTGTGAAAAAGCGGTGAAAGCTGGGGGGCCAGAGAGCTGG 3859  
QY 1103 ----- 1103  
Db 3860 GGGAAAGGGCAGAGAGACTGTGGGCGCAGAGAGAGGCGAGGCCCTTCACCCC 3919  
QY 1103 ----- 1103  
Db 3920 CATCAAGATATTCTGAGATCTACACAGCTTTGGCAGAGCTGGGAGCCACGCGAGACAC 3979  
QY 1103 ----- 1103  
Db 3980 AGGGCTGCCGCAACCCAGGGGTGAGACCTGTAGCTCTTCTGGCCCCCTTTTGCTCC 4039  
QY 1103 ----- 1103  
Db 4040 TGGCTGAGCTGCAGAAATCGGCTGACAGGACTTTGTCTAGCGCATTTCTTCCCTCCAGG 4099  
QY 1103 ----- 1103  
Db 4100 CAGAAAGAGCTGGCTGGCTCCCCCTTCAGAGCTCCACAGAAAGTAGAGGCCAGAGCTC 4159  
QY 1103 ----- 1103  
Db 4160 ATTGTCTGTCCAAACAGACCCCACTTTCAGAGTCCGTGTTCGGGTCTGGCTTGGAGAG 4219  
QY 1103 ----- 1103  
Db 4220 GGGGTGAAATGAAGCTGGGGGGCAGGCTGAGGCTCGTGGGCCCCCTTGGCACTCATG 4279  
QY 1103 ----- 1103  
Db 4280 GGTGTCTTAGGGTGCAGGCTTCTTCTCTGCTAGCTGTGACAGCCCAAGCTCCAG 4339  
QY 1103 ----- 1103  
Db 4340 GAATGAGTGTAGGAGCAAGAGCAGGCCAGTGAAGTGAAGCTTTAGGGCAGAGACCC 4399  
QY 1103 ----- 1103  
Db 4400 CAAGGACCTTATCTGAGAGAGAAAGAGTCCAGGGCGTGGAGGCTGGCGGAAACC 4459  
QY 1103 ----- 1103

Db 4460 AGGCTGGGGCATTCCTGCTCTTTGGCCAGCACTCAGGCTCAGACTCCTCGAGCAGCAG 4519  
QY 1103 ----- 1103  
Db 4520 GGGTGGGGAGAGAGTGGAGGGGGTGGTGAAGATATTCTGCTTTTATTCAGCTG 4579  
QY 1103 ----- 1103  
Db 4580 TGGGAGCCAGAGGTAGCAGCAGGAGAAAGCAACCCAAATTGCTGAGTCCCTCC 4639  
QY 1103 ----- 1103  
Db 4640 AAGCTGGGAACAAGTGAAGACCTCATAGAGGGATGCTGTAGCTGACCCAGGGGTGG 4699  
QY 1103 ----- 1103  
Db 4700 ATGAGTCAATCCCTTAGGAGCCTTCAATTCATCAGGAGGAGGAAAGAAAGA 4759  
QY 1103 ----- 1103  
Db 4760 GGCTGGCAGGCCCTTACTCCACCCAGGAGTGTGGCTGAAGGGGAAAGAGGGTATG 4819  
QY 1103 ----- 1103  
Db 4820 CAGAGTCAAGGGGTGGCTCACTCAGGGTCTTGTGAGCAGATGCTCCCTGCCCCAC 4879  
QY 1103 ----- 1103  
Db 4880 CTTCAAGTAGCCACAGAAACAGAGAGAGGCTTTCAAGGACTGTGCTGAGTT 4939  
QY 1103 ----- 1103  
Db 4940 CTGAGCTCTGAGACTTGAGTTCCCTCCAGAGAGCGGAGAGAGGCTGAGACTCC 4999  
QY 1103 ----- 1103  
Db 5000 TGTGTCTCAACTGTCTGGGCACTTGAGACAAAGTGGAAAGAGTGGAGAGAGGCC 5059  
QY 1103 ----- 1103  
Db 5060 TCTCCCTGTGTCTTCACACCCAGGCTGCTTGTGAGCGTGGCCAGGACTGGCAGA 5119  
QY 1103 ----- 1103  
Db 5120 GAAGCGCCCCATTAAAGCTTGGGTCTCCACAGGGCAGAGGCTTAGACTTGTGCC 5179  
QY 1103 ----- 1103  
Db 5180 TAGTCTAGAAAGCTCGGGGCAACATCCAGACGGCTCCTCTGGCCCCGGAGCTTGG 5239  
QY 1103 ----- 1103  
Db 5240 CCCCACTCTCTCCCTTGGGCTGACCTTGGAGGAGTCAACAGGCAACCCCAAGTT 5299  
QY 1103 ----- 1103  
Db 5300 CAGGAGCGCGGGGTAGGGGAGTGAATCCACTGCCCCCTTTGTTTCTCTCTTGT 5359  
QY 1103 ----- 1103  
Db 5360 TCTTGTCTCTCTTCTTCTTCCGCCCCGCGCATGAGAGCCCTGACAGACTGCTTC 5419  
QY 1103 ----- 1103  
Db 5420 TCCGATCCCTCTCCCTGCTCTGCGGCTCTCGGCTTTCACATTGCCATTGCGCT 5479  
QY 1103 ----- 1103  
Db 5480 CCTCTCTGAGGCTTGTCTTCCGCTTGTGATCTCGATTTGTGCTCACTCTCTTC 5539  
QY 1103 ----- 1103



Db 5540 CCTATGCTGCACTGCCCCAGCCTGGCCGATGCTCCCTGAGGCTGAATTCCTC 5599  
QY 1103 ----- 1103  
Db 5600 GCGCTTTCTTCTCTGTGCTAGACACGACAGCTCCTGTCCATGCTGTGGGAGCTCCC 5659  
QY 1103 ----- 1103  
Db 5660 AGAAGACAGAGCCCTCTTGAGCTCCAGTCTCCGCTGGACATCTGCCAGGGAGAGAG 5719  
QY 1103 ----- 1103  
Db 5720 GCTGGAGGCTGACCGGAGCCTGGACCTTTCCAGTGTGCTGGCTAAGCTTGGACAC 5779  
QY 1103 ----- 1103  
Db 5780 AGGTGCAATTCCTGGGGTCAAGAGCCATCCAGCTGTCTTTCTGTCCAGAGAGACA 5839  
QY 1103 ----- 1103  
Db 5840 GCTCAGGCAAGTACTCCCAACCCCTGTCCCATCTGCTTTGCACTGAGGGTGACC 5899  
QY 1103 ----- 1103  
Db 5900 TAGGTCAAGGCAAGGGAGGGTGATATGAAAGAGAGACAGTGGTTGACTTTGGGGA 5959  
QY 1103 ----- 1103  
Db 5960 TGGGGGACAACAGAGAGGGGGGAGTGGCCAGATGTGTGACAGATGGGAAGACA 6019  
QY 1103 ----- 1103  
Db 6020 GAGACATGCTGGAGAGGGGCCAGCGGGCTTAAGACATGCGGTGGGGGACACAGG 6079  
QY 1103 ----- 1103  
Db 6080 GAGAAGTGTGTTCCCAAGGCTGAGGGGTGGCAAGTTTCCCGCTCATGGCTTGG 6139  
QY 1103 ----- 1103  
Db 6140 TTGCCCCCTCTCTCTTATATAGTTATGAGCACTGTGTGTGAGTCTTGGGCTGCTT 6199  
QY 1103 ----- 1103  
Db 6200 CTGCTGAGAGGAATATCCGTCAGAGATCATTAGGAGATATACAGATTAACAAG 6259  
QY 1103 ----- 1103  
Db 6260 CTCTGTGCACTGAGCATCTGTGTGCGGGGAAGTCTGAAACGAGAGAAAGTAGACC 6319  
QY 1103 ----- 1103  
Db 6320 TGATGACATAATCTGGGAGGCTTCTGAGCAGGTGAGTAAAGGAGGAGTTAAGACT 6379  
QY 1103 ----- 1103  
Db 6380 CAGAGACTGTGGCTGTGTCTGGGTGGAGAGACAGCTGTGACAGCACTTTGAGAC 6439  
QY 1103 ----- 1103  
Db 6440 GAGAGCAGAGAGAACAGAGTGGGTGATGACAGAGGCTTGTGTGAGAGCTAGTAG 6499  
QY 1103 ----- 1103  
Db 6500 AATAAGCAATGATAGGCTGAGGGTATGTTGGGCACTGAGTGAAGACATGCTGAGAT 6559  
QY 1103 ----- 1103  
Db 6560 GGGGGTCTCTTTAGACCTTCCGGCTTCAGCTAGCACCGCTGTATCACTCAAGCAC 6619  
QY 1103 ----- 1103  
Db 6620 GCAAGTAGTACACAACAGATGAGAAATTGTTCCATGTCCCTTGCTACTCCACCCC 6679

QY 1103 ----- 1103  
Db 6680 ACCCAGAAAACATGGGCTCCTGCCGTTACTGTAGGGTAAGCATTTGTTTATATC 6739  
QY 1103 ----- 1103  
Db 6740 TAGTGGGCTCCCCCTGATATGTGTTTGTGTTGTTGTTGGAGAGAGTCACTTG 6799  
QY 1103 ----- 1103  
Db 6800 TCAACCAAGGTGATGACATGAGCGGATCTCGGCTCATGCAACCTCTGCTGGGT 6859  
QY 1103 ----- 1103  
Db 6860 TCAAGCAATTCCTGCTCAGCCTCTGATATCTGGAGCTCTGATAGAGCGGGTTTAC 6919  
QY 1103 ----- 1103  
Db 6920 ACAG3ACCCACACACACGCCCACTAATTTTGTATTTTATAGAGACGGGGTTTAC 6979  
QY 1103 ----- 1103  
Db 6980 CATGTTGGTCACTGATCTGAATCTGAACTCTGACCTGTGATCCACCGGCTCAGCTCCA 7039  
QY 1103 ----- 1103  
Db 7040 AAGTTGGGACTACAGGGGTGAGCCATGCGCCGCCCTGTATGTTTAAAGTCC 7099  
QY 1103 ----- 1103  
Db 7100 GTTATCTTTGTGAGACTGATATGATTTGATCCATCTCAGTGTCTTGAACACC 7159  
QY 1103 ----- 1103  
Db 7160 TACTATGTGAAGCATTTGTAATTAATACAAATATGCAATGCACTAGCTATGCTGT 7219  
QY 1103 ----- 1103  
Db 7220 AATCCAGTCTTAGAAGGCTGAGATGAGATCACTTGAGGCCAGAAATTGAGACCA 7279  
QY 1103 ----- 1103  
Db 7280 GCGTCGACACACAGAGAGCCCATCTTACACACAAACAAATTAAGCAGCATAGT 7339  
QY 1103 ----- 1103  
Db 7340 TGTAGCACTGTGTGTAGCTATTTGGAGGTTGAGGTGGAGAGATGCTTGAAGCCA 7399  
QY 1103 ----- 1103  
Db 7400 GAGTTGAGGCTACAGTGAATGATGATGTCGCTCTCAGTGTAGGAGACAGAG 7459  
QY 1103 ----- 1103  
Db 7460 AGTGTCTCTTAAAAAAGAAAAAAGACTATCCCTTCTTGCCCTGTAAAGCAATTT 7519  
QY 1103 ----- 1103  
Db 7520 CTTCTTCTCAGGCCCACTTCATTTTATAGCTAATAATAATCATTTCCATGTGTA 7579  
QY 1103 ----- 1103  
Db 7580 TGTGATTCAGTGTATATGTAATTTCTGGCTTAAGGGTGTGTGGATGAGACACC 7639  
QY 1103 ----- 1103  
Db 7640 AGCAATGTATTTAGTGTCTTACTGTATTAAGTGAATGTGTTGCTGGGGCATAGCA 7699  
QY 1103 ----- 1103  
Db 7700 GGCATTTGGGTATATGTTGAGCCAAGTGGATGATGATGAGAGGCTTAGACCA 7759

QY	1103	-----	1103
Db	7760	GAGAAAGACTGACTCAAGTGGGGCTCTGGGGTCTCAGATGGAGACTGCCAGGCCCTT	7819
QY	1103	-----	1103
Db	7820	CTATCTGCCCTTCCGGTCCCAACCCGACTCACCCTGTCTCCAGCATGGGGCAGTGGG	7879
QY	1103	-----	1103
Db	7880	TGATGATGTACTGGTTAAGTGGCTCAGCTGATTTCTGGGGTGGGGCCCCCAGGCT	7939
QY	1103	-----	1103
Db	7940	ACCCACAGACGCCAGAGAGTTGGCGGGAAAGATGCTTTGTGTGATGTTCATGGG	7999
QY	1103	-----	1103
Db	8000	GATGACAGGTCTGGCGGCCCAAGCAGCATGGGGCTCAGTGGAGGGGACTAGTGGCC	8059
QY	1103	-----	1103
Db	8060	CACGTACCCAGTTTACCCACTCCCCCAGCCCCCGCAACTCCCTGCTGTGTGGGGAG	8119
QY	1103	-----	1103
Db	8120	TGGGAAACAGACCATGGTTCGATTATCGTGTGTGAGGTGTAAGGCTGAGGGGCA	8179
QY	1103	-----	1103
Db	8180	GGGGGATGATTTTGTGGGTCTCAGGATGCTGTGGCCCTGTGAAAAATGTAAGAAAA	8239
QY	1103	-----	1103
Db	8240	CAACATAGGGTTCGCCGCGAGGACAGCAAGACTCTGGGGGAAAGGAGCAGAGCGGTGC	8299
QY	1103	-----	1103
Db	8300	TTGGGCAAGGAAAGGTGCACAAACATTAAAGATGTGCGGTGCATCAGCGCAGAAAGTT	8359
QY	1103	-----	1103
Db	8360	TTGGAAAGATGCAMAGAGAACTGCAAATCTCATCCCTGAGCCCTGAAAGATGATTG	8419
QY	1103	-----	1103
Db	8420	GCCCAACAGTGAATACCTGGCTCAGCACCTTATTCAGATTGAAAAGAGAGCCTTCT	8479
QY	1103	-----	1103
Db	8480	TGTGGGTGGTGCAGGAATATGCAAGGCTGAATTGAGCCCGAGGCTGGGTCTGACTGTG	8539
QY	1104	-----	1119
Db	8540	TCCCTCTCCACAGACGGGCTACACAGCTGACCCCGCAGTCCGGCTGCTGATCCATC	8599
QY	1120	ILethrAepGIumeLethrHisSerILethrLeuArgLeuGluApmEserProGlu	1139
Db	8600	ATCACCGATGAGTACTCACCCACAGCATCAGCTGCGCTGGAGGACATCTCACCCGAG	8659
QY	1140	ArgPheLeuSerProLeuLeuGlyLeuPheILegInaValAlaILThrLeuAlaThr	1159
Db	8660	CGCTTCCTGTACCACTGCTAGGCTCTTCATCCAGGGGTGGCGCCACGCTGGCAAG	8719
QY	1160	ProProAepHISValValILPheAenValGlnArgAepThrAepAlaProGlyGlyHis	1179
Db	8720	CCACCGGACCAAGTGTGTCTTCAACCTACAGGGGACACCGCAGCGCCCCGGGGGCGAC	8779
QY	1180	ILeLeuAenValISerLeuSerValGlyILnProProGlyProGlyGlyILProProPhe	1199
Db	8780	ATCTCTCAACGTGAGGCTGTCCGTGGGCGACCGCCAGGCGCGGGGGGGCGCCCTTC	8839
QY	1200	LeuProSerGluAepLeuGlnGluArgLeuTyILeuanArgSerLeuLeuThrAlaIle	1219

Db	8840	CTGCCCCGTGGAGAACCTGCACAGAGCGCCTTAATCCCAACCACGACTTCACAGCGCATC	8899
Oy	1220	SerAlagInAryValIleuProPheaspAspaeniIeCySleuArgGl.uProCyeglunsn	1239
Db	8900	TGCGCACGCCGCGTCTCCCTTGACGACAACAATTCTGCTCGCGGAAACCCCTGCGAGAAC	8955
Oy	1240	TyrMetArCy.sValSerValIleuArgPheasPerSerAlaproPheIlalaserSer	1255
Db	8960	TACATGCGCTCGTGTGCGTCTGCGCTTCGACTTCCTCCGCGCCCTTCATCGCCTCTCC	9015
Oy	1260	SerValIleuPheaArgProIIehiAProValGelyleuAlyCy.sArgCy.sProProgly	1275
Db	9020	TCCGGCTCTTCGCGCCCATCCACCCGTCGGAAGGCGCTGCGCTCCGCTCCGCGCGC	9079
Oy	1280	PheThrGIyaSPYrCy.sgluThrgluValAspleuCy.sYrSerArgProCy.sgl.yPro	1299
Db	9080	TTCAcGGGTGACTACTGCGAGAACCGAGGTGAACTCTGTCTACTGCGGCCCCTGTGGCCCC	9139
Oy	1300	HiselYarCy.sArgSerArggluGelyleYrThrCy.sleuCy.sArgAspgl.YrThr	1319
Db	9140	CACGGCGCTGCCGACGCGGAGGGCGGCTACACTGCTCTGTGCGATGGCTTACAG	9189
Oy	1319	-----	1319
Db	9200	GGTAGCCMAAGGAGGGAATCATGAGGCGACGCTTGAAAGGTGACTGTGTGTCAGGC	9259
Oy	1319	-----	1319
Db	9260	ACAANTCAGACAANAATGCTGGCGGCTGCCTCATCTCTTCCGAGTAGAGTGACATCA	9319
Oy	1319	-----	1319
Db	9320	CTGAGAGGTCAATAATCCGCGCTGCTCCTGACAGCAACCCACTGAGGAAGGAGTGCAA	9379
Oy	1319	-----	1319
Db	9380	GGAATGCTTCTCCCTGCTGTGACAGACACACAGGCAGTGCCAATTGTGGCAGAAGCTGG	9439
Oy	1319	-----	1319
Db	9440	CTCCAGAGCAGTGGCGCTGTGTGCTGTATCAGCTTTGCGCGTAAGTAGCTGTGACT	9499
Oy	1319	-----	1319
Db	9500	TGAACAAGTTACTTAACCTTCTGCGCCCTCGTTATTGGAANTGAGATATGACCGTAC	9559
Oy	1319	-----	1319
Db	9560	TTAGGTCATAAGTTGGTTGTGAGGATTAATGTATTAATTGATATTGAAAACAGCTGG	9619
Oy	1319	-----	1319
Db	9620	TGGATGGAAGGTGTGGCTTGATTTAAGGAAACGTAGATTGGNGACTTGCTCTG	9679
Oy	1319	-----	1319
Db	9680	GTCCTCATCACACCGGTTTAGAGGTGAGGCTGGGCTAAGAACTGTGCTCTCCCCACTG	9739
Oy	1319	-----	1319
Db	9740	AGCAGTTCCTTTCTAGAGCTGACAGCAATCTTTCTCCCAAGCATTAAAGGAAAGAC	9799
Oy	1319	-----	1319
Db	9800	ACAGAGAAGCTGGGGAACTGAGAGGGGTGAGTTCTGAGATGGGTGTCTCAGAAAAAC	9859
Oy	1319	-----	1319
Db	9860	ATTGCACCTGTATGGGCTTGGGCAGAATGGCTCCGCGCCAGCTGCTGGTTAAAGA	9919
Oy	1319	-----	1319

Db 9920 GGAATCTGGCTAAGCGGCTTCCTGTGGCTGTCAAGCCAAAGTCTTTACACATCTGTCT 9979  
 Qy 1319 ----- 1319  
 Db 9980 CCGTGGAGTGTCACTGTCTCTGTCTATAGTCTCTGACCCCTCCCTTCAAGGCTCTGTGT 10039  
 Qy 1319 ----- 1319  
 Db 10040 CTTCCTCAGAACACCCCTGTCTCCCTTCCCTGTGGGCTTTCTTCTAGATGCCAGGC 10099  
 Qy 1319 ----- 1319  
 Db 10100 CCTGTCTGTGTCTCTCAAGGCGCATGGCCCACTGTCTGTCCAGACGCATAGAGAGGCC 10159  
 Qy 1319 ----- 1319  
 Db 10160 CACTCTTCCCCGCTGATCTTCTGCCCCAAGACTTGGCAGACTTGTCTCTGTCTG 10219  
 Qy 1319 ----- 1319  
 Db 10220 CTGCTAAGGGGTGGGGGAGAGGAGGACGAGGCTTGGGGGAGGAGAGAGTCTTGTCTC 10279  
 Qy 1319 ----- 1319  
 Db 10280 CAGAGGATGTGGGGGGGTTGGAGGGGGGGTTAGACAGCGCAACACAGTTTCC 10339  
 Qy 1319 ----- 1319  
 Db 10340 ATGGAACACAGAGCGACGTACTAGGCACAGAGCTGAGAACCTATCTTAGAATCTGGGA 10399  
 Qy 1319 ----- 1319  
 Db 10400 ATGGGAGGAGAGAGAGGTGATTTTCCCATCTGATGTGGATTTTCAGCTCTCTGT 10459  
 Qy 1319 ----- 1319  
 Db 10460 TGAATCTTGTCTGTCTCAAGGCTCAGTTTCCCATCTCAGCAATGCCAGGTGCACAG 10519  
 Qy 1319 ----- 1319  
 Db 10520 AGGTCTTATGTGTGACCGCAGGCTCAGAGTGGAGTCTCTGTCCCATTTCTGGGGGC 10579  
 Qy 1319 ----- 1319  
 Db 10580 TCCCCAGTCTCTCTCATACTTTCATTAACCCAGCAAACTAATGAGAGCTGTCTC 10639  
 Qy 1319 ----- 1319  
 Db 10640 CCATCCCCAGGAGGAACTGATCTCAGTGTCTGAGAGGAGCAAAAGCACAGTGGGAA 10699  
 Qy 1319 ----- 1319  
 Db 10700 GGAAGGAGAGGAGGCTCCAGGCTCCGCTGTCTGTGAGGCCATCTGTGTGTGAC 10759  
 Qy 1319 ----- 1319  
 Db 10760 ATTATCCCTTGTGGGGCCAGAGACCCCCAGGTGTGTGTGTGATGACATCGAAGGC 10819  
 Qy 1319 ----- 1319  
 Db 10820 TCCCGGATCCGGGCTTTACACCAAGTGTCCAGCAGGCTGTGTGTGAAACCCAAA 10879  
 Qy 1319 ----- 1319  
 Db 10880 TGGGTAGGAAAGTGTGTGGGACTCCAGAGAACTTGACAGGAGGAGCAGTGTGAGCTG 10939  
 Qy 1319 ----- 1319  
 Db 10940 TGGGAGAGCGCTCAGTGAACCAACAGAGGTGCCATGAAGGCTGGGAGCCATAGGTC 10999  
 Qy 1319 ----- 1319  
 Db 11000 TCAGAGAAATGGACACACGAGAGAGAGTGGGATGCTCTGGAGGTTTGGGAGATTAAG 11059

Qy 1319 ----- 1319  
 Db 11060 AGGTACACGGAGAGCCATGGAGCTATCACTGGGTTGTGGGGGTCTCAGTTCCCTCT 11119  
 Qy 1320 ----- 1325  
 Db 11120 GTCTCAGGACTTGTGACTACCTGTCTCTTGTCTGTTCAGAGTGAAGCACTGTGAGTGC 11179  
 Qy 1326 SerAlaArgSerGlyArgCysThrProGlyValCysLeuAspGlyGlyThrCysValAsn 1345  
 Db 11180 AGTGTGCTCAGAGCGGTGCACCCGGGTGTGTGAAGAAATGGGGGCACCTGTGTCAAC 11239  
 Qy 1346 LeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAspPheGlyLeuValGln 1365  
 Db 11240 CTGTGTGGGGCGGTTTCAAGTGCATGCCATCTGAGAGACTTGAGAAAGCCCTACTGC 11299  
 Qy 1366 GlnValThrThrArgSerPheProAlaHisSerPheIleThrPheArgGlyLeuValGln 1385  
 Db 11300 CAGTGACACACGCGCAGCTTCCCGCCCACTCTCATCACTTTCGCGGCTGTGCCAG 11359  
 Qy 1386 ArgPheHisPheThrLeuAlaLeuSer----- 1394  
 Db 11360 CGTTTCACTTCACTGAGCCCTCTC-GTGAAGTGTGGGCACTGGGGGTGGGAGTGGG 11418  
 Qy 1394 ----- 1394  
 Db 11419 CCGTGTGGGCACTGAGGTGCTCTCTGTGTGTGTAAGTAAGAGTCAAGGCAGTGA 11478  
 Qy 1394 ----- 1394  
 Db 11479 AGCGTGACAGAGCTGCCGAGACAGCAGTCTGTCTTGAGATTGCTGTGTCTGTGCC 11538  
 Qy 1394 ----- 1394  
 Db 11539 AGCAGATCTCCCTCCCTGCGCTGTGTAGGTGGGGGTCTGGGATCCAGGTGGGTAC 11598  
 Qy 1395 ----- 1402  
 Db 11599 CCCATTCCTGCCCCCATCCCAACTCTGTTCAGGTTTGCCCAAAAGAGCGGACGCG 11658  
 Qy 1403 LeuLeuLeuTyrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeuGlyValIle 1422  
 Db 11659 TTGCTGTGTACAAATGGGGTTCATATGAGAACATGACTTGTGTGCTCTGAGGTGATC 11718  
 Qy 1423 GlnGlnGlnValGlnLeuThrPheSerAla----- 1432  
 Db 11719 CAGGACAGGTCTCAGCTTCACTTCTGTGC-AGGTATCAAGTGTCCCCCATCTTGGCC 11777  
 Qy 1432 ----- 1432  
 Db 11778 CATCTTCAAAAGGCCCAAGTCTTTCAGGCCCTGACCCCAAGCACATACTATCAGCC 11837  
 Qy 1432 ----- 1432  
 Db 11838 AAATCTGGGCCAGCCCAAGCCCACTGGCACTGGCAACCCCTGCATCTTCAAGCCCTGTCA 11897  
 Qy 1433 -----GlyGlySerThrThrValSerProPhe 1442  
 Db 11898 GGCATTCACGCTCACCTGATCTTTCCTCCCAAGGAGTCAACACACGAGTGTCCCATTC 11957  
 Qy 1443 ValProGlyGlyValSerAspGlyGlnThrPheThrValGlnLeuLysTyrTyrAsnLys 1462  
 Db 11958 GTGCCCGGAGAGTCAGTATGCCAGGTGCATACGCTCAGCTGAAATTAACAATAAG 12017  
 Qy 1462 ----- 1462  
 Db 12018 GTGGGTGGAGGGACACAGAGGTTGGGGTTCGTGTCTTTCCTCAAGGTCTTACCAGC 12077  
 Qy 1462 ----- 1462  
 Db 12078 CCGAGTGGCAATTGCTCAGGCTTGGGTGGGTGTGAGGCAATTCTGTGAGAGAGA 12137





QY 2004 ----- 2004  
 Db 16517 GACTGTTGTTGTTGTTGTTGTTGTTTCTTGAGACAGAGTCGTGACACC 16576  
 QY 2004 ----- 2004  
 Db 16577 CAGGCTGTGTGAGTGAGCGGTGATCTCGCTGCTGCAACTCTGCTCCAGGTTCAAG 16636  
 QY 2004 ----- 2004  
 Db 16637 CAATTTTCTGAGCTTGTGGCGGCAACCACTGGCTAATTTTTTTTTTTTTT 16696  
 QY 2004 ----- 2004  
 Db 16697 TTTTGTGATTTTGTAGAGACAGGTTTTCACATGTTGGCAGTCTGTCTGAA 16756  
 QY 2004 ----- 2004  
 Db 16757 CTCTGACCTCATGATCCGCTGCTGCGCTCCCAAGTGTGGATTAACAGCGTGAG 16816  
 QY 2004 ----- 2004  
 Db 16817 CCACTGACCCCAACACATTTTGGACTTTTTTTTTTTTTTTGAGAGAGTCTGCTC 16876  
 QY 2004 ----- 2004  
 Db 16877 TGTTCATCAGGCTACATGACATGAGTCATGATCTCGCTCTGCAACTCACTCAGAG 16936  
 QY 2004 ----- 2004  
 Db 16937 TTCAAGGATTTCTCTACCTCAGCCTTCCAGATGCTGTGATCAAGCGCTGTGATCA 16996  
 QY 2004 ----- 2004  
 Db 16997 TGCCAGCTAATTTTTATATATTTTTTATAGAGACAGGTTTTCACATGTTGGCCAGA 17056  
 QY 2004 ----- 2004  
 Db 17057 TGGTCTCGATCTTGAACCTGTGATCAGCCGCTGCGCTCCCAAGTCCGAGATTA 17116  
 QY 2004 ----- 2004  
 Db 17117 CAGCGGTGACCAACCGTCCAGCCACATCTTGGACTTTTACAAATAATGCTGAATAAC 17176  
 QY 2004 ----- 2004  
 Db 17177 CTGCACTGTGCTTGGCATAGGGGCTGTGTAAGATTCAAAAGATGATGACAGGTTCC 17236  
 QY 2004 ----- 2004  
 Db 17237 AATCGTCAGCTACACCGAGTCTGTGATTAAGCCACCAAAAGCAGACAGTCTGGGG 17296  
 QY 2004 ----- 2004  
 Db 17297 AGTAGGGGGTGTGGCTGTGACCCAGAGACACCAAGCCAGATGGGGGCTGGCGCTGGT 17356  
 QY 2004 ----- 2004  
 Db 17357 CTGGACTGGAATGAGCAGGTTTCATAGCAGAGAAAGTGGGAGAGCAAAAGTGTGC 17416  
 QY 2004 ----- 2004  
 Db 17417 TGAAGACAGCAGAGATGGGGCTCAAGAGATGTATGCTGTTCCTTAGAGCCCATG 17476  
 QY 2004 ----- 2004  
 Db 17477 TGTGATCTGATCTTGGCGAGATTTTGGCCCTGTGTCTTACATACCTGTCTCTGTTT 17536  
 QY 2004 ----- 2004  
 Db 17537 CACGAGGCCCTGTGTGTAGTTACTGTCCCGGCCCATGAGGTGTGCTCAGCAG 17596  
 QY 2004 ----- 2004

Db 17597 AACCTCTCAGCAGTCTCAGGGGCTTCTGTGTGTCTCCCTGAGGGCGGCTTTGGCT 17656  
 QY 2005 -----  
 Db 17657 TCCCTCCCGGAGGAGCTGTGTGGCGCACTGTATAGCAAGAGGGGTGCTCCCCCA 17716  
 QY 2021 snleupheanCystrSer1LeThrPheSer1leuLeuysglYpHe----- 2036  
 Db 17717 ACCTTTCACTGACGTCCATCACTTCTTCAGAACTGAAGGGCTTGTATGAACCCC 17776  
 QY 2036 ----- 2036  
 Db 17777 CTGATCTCATCTTTTCCCTGTCTTGTCTGATGATCTGACCTGCCCCACCTCCATCTT 17836  
 QY 2036 ----- 2036  
 Db 17837 TGAAGACGGGGCTTGTGAATTCAGCCTGTGTGTCTGTGGGCTCCAGCTGAGAGGCC 17896  
 QY 2036 ----- 2036  
 Db 17897 GTCTCACTCTGAGTCTTTTGTCTCAGAGTTCTGCTCTCACTCTGCTCTCCG 17956  
 QY 2037 -----  
 Db 17957 CTCCGTCTGTCTTGTGAGCTGAGCGGCTACAGCGAATGATGAGGCTGACGAGGCGC 18016  
 QY 2052 SerGlnGlnleuAlaLeuLeuLeuArgAlaThrGlnHisPheThrGln 2071  
 Db 18017 TCCAGCAGCTAGCCCTGCTCTGCGCAAGCCACACACACACACACACTGCTACTTCGGC 18076  
 QY 2072 SerAspValIysValAlaTyrGlnleuAlaThrArgLeuLeuAlaHisGlnSerThrGln 2091  
 Db 18077 AGCGACGTCAAGGTGCTTACCAAGCTGGCAAGCGGCTGTGCTGCCACAGAGCAACCA 18136  
 QY 2092 ArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGln----- 2106  
 Db 18137 CGGGGCTTGGGTGTGTCTGCCACACAGAACGTGCACTGAGGTGGGGCTTGGAGGA 18196  
 QY 2106 ----- 2106  
 Db 18197 TGAAGGCTGGCTGTGATAGTATAGGGGTCAATGTGATGAGCTGTCAATGGCAACTGGG 18256  
 QY 2106 ----- 2106  
 Db 18257 GGGCAGAGGGGCGCTCCATCCACCTACAGAGAGCTCCGTGTGCGAGAGGCTCCA 18316  
 QY 2106 ----- 2106  
 Db 18317 TGAACCTGTGACCCCTGCGCCAGCCCTTCCCAACCTTCTCATATGATATGTG 18376  
 QY 2106 ----- 2106  
 Db 18377 TGTGGGGGATGTGTGACCCAGAGCAGAGGCTGTGCTGTGTGATCCATGGGCTGGAGGA 18436  
 QY 2106 ----- 2106  
 Db 18437 AGCATATCTCAGACATGTGTGTGGGTCCAGGCAAGGGCTGGAAAGCTTATATGTA 18496  
 QY 2106 ----- 2106  
 Db 18497 GAATGAGAGAGGGCTTAAGGGCAGAGCATCACTGACCTGATGTGGCAGTGGCTGCA 18556  
 QY 2106 ----- 2106  
 Db 18557 GGAACATTTCAAGGGAGGCAAGTGAAGCAGAGAGAGGGGCTGTGTGGAGAGCAGGGA 18616  
 QY 2106 ----- 2106  
 Db 18617 CTGCCCCGGGCAAGCGGGGCTCCATGTGTGGGATGTCAGGTGTGGTGTGAGACA 18676  
 QY 2107 -----  
 AsnLeuLeuArgValGlySerAlaLeuLeuAspPh 2118

Db 18677 CCCACCGTGAACCTTGCCCAACCCCAAGAAATCTGCTGGGGTGGCAGCGCCCTCTGACAC 18736  
 Qy 2118 TAAAsnLysArgHisTrpGluLeuIleGlnGlnThrGluGluYsLysThrAlaTrpLeu 2138  
 Db 18737 AGCCAAACAAGGGGCACTGGGAGCTGATCCAGACAAGAGGGTGGCAACCGCTGGCTGCT 18796  
 Qy 2138 uGlnHisTrpGluAlaTrpAlaSerAlaLeuAlaGlnAsnMetArgHisTrpTyrLeuSe 2158  
 Db 18797 CCAAGACTATGAGGCTTACAGCCAGCTAGCTGGCCCTGACAAATCGGGCACAACCTACCTAAG 18856  
 Qy 2158 rProPheThrIleValThrProAsnIle----- 2167  
 Db 18857 CCCCTTCAACATGCTCAAGCCCAACATTTGTAAGGCTGGTGCCTGGGTTGGGAGGGGTT 18916  
 Qy 2167 ----- 2167  
 Db 18917 TGTGAGGAGAGTCCCGCAAGAGGGGCTGTGCTGGGCTCTGCTCAGAGTCCG 18976  
 Qy 2168 -----ValIleSerValValArgLeuAspLysGlyAsnPheA 2180  
 Db 18977 ATCTGTGACCATCCCTCTCTTAAGTCATCTCCGTAGTCCGCTTGACAAAGGGAACTTTG 19036  
 Qy 2180 IAGIYAlaLysLeuProArgTrpArgIuaIaLeuArgGlyGluGlnProProAspLeuGluT 2200  
 Db 19037 CTGGGGCCAAAGCTGCCCGCTACAGAGCCCTGCGTGGGAGCAGCCCGGACCTTGAGA 19096  
 Qy 2200 hTrpValIleLeuProGluSerValPheArg----- 2210  
 Db 19097 CAACAGTCATCTCTCGTAGTCTGTCTTCAG-AGGTCAGTGTGGCCATGGAATTGAGTTG 19155  
 Qy 2210 ----- 2210  
 Db 19156 GGAAGCTGGAACCCAGTGTCTGTGAGACTCCACAAGAGCAGGGCCAGCTTAAGTGTGAC 19215  
 Qy 2211 -----GluThrProProValValArgProAlaGlyProGlyGluIaGlnG 2226  
 Db 19216 AGTGTCCCTCCCAAGAGAGCGCCCGGTGTCAAGCCCGCAGGCCCCGAGAGAGCCCAAG 19275  
 Qy 2226 IuProGluGluLeuAlaArgArgIuaArgArgHisProGluLeuSerGlnGlyGluIaV 2246  
 Db 19276 AGCCAGAGAGAGCTGGCAGCGGAGCAGCGCACCGGAGCTGAGCCAGGGTGAAGCTG 19335  
 Qy 2246 aAlaIaSerValIleIleIleTrpArgTrpThrLeuAlaGlyLeuLeuProHisAsnTyrAspProA 2266  
 Db 19336 TGGGCAAGGTCAATCATCTTACCGCACCCCTGGCGGCTACCTGCTCATTAATGACCTTG 19395  
 Qy 2266 sPlyAspSerLeu----- 2270  
 Db 19396 ACAAGCCGACGTTAGGTGACAGACTAGGGGACAGGTGTGGGTATGGGTCCGGC 19455  
 Qy 2270 ----- 2270  
 Db 19456 GGTGAGTGTGAGGACATGAGAGGGGTCCGGGGGCGTCTCCCAAGTCATGTGACTCCGTGG 19515  
 Qy 2271 -----ArgValProLysArgProIleIleAsnThrPro 2281  
 Db 19516 TGAAGTGTGACCTGACCTGACCCCAATCAGAGTCCCAACCGCCGATCATCAACAACCC 19575  
 Qy 2282 ValValSerIleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysPro 2301  
 Db 19576 GTGGTGAAGCATCAGCGTTCATGATGATGAGAGGCTTCTGCCCGGGCCCTGGAACAAACC 19635  
 Qy 2302 ValThrValGlnPheArgLeuLeuGluIuThrGluIuaArgTrpLysProIleCysValPhe 2321  
 Db 19636 GTCAAGGGGCAAGTTCCCGCTGCTGAGACAGAGAGCGGACCAAGCCCATCTGTGTCTTC 19695  
 Qy 2322 TrpAsnHisSerIleLeu----- 2327  
 Db 19696 TGGAAACCATTAATCTGTGAGCTGACATGCCCTCGCCCTCAGAGCTTCGGGCTGAAA 19755  
 Qy 2327 ----- 2327  
 Db 19756 GTCCAGAGCCCTGCAATGCTCAACCTCTGTCTCCCTGACCCAGCTTCCTCAACAGAG 19815

Qy 2328 ValSerGlyThrGlyGlyTrpSerAlaArgGlyCysGluValValPheArgAsnGluSer 2347  
 Db 19816 GTCAAGTGGGACAGGTGGCTGTGGTGGCCAGAGGCTGTGAAGTGTCTTCGCAATGAGAGC 19875  
 Qy 2348 HisValSerCysGlnCysAsnHisSmetThrSerPheAlaValLeuMetAspLysArg 2367  
 Db 19876 CAGTCAAGCTGGCAGTGGCAACCAATGACGAGCTTGCTGTCTCATGACGTTTCTCGG 19935  
 Qy 2368 ArgGlu----- 2369  
 Db 19936 CGGAGAGTGGGGCCCAAGGGGAGCTGACAGAGCCGTGGGTGGGACCCAGGGCAGGGG 19995  
 Qy 2370 -----As 2370  
 Db 19996 CTGGGTGTCAAGGTCTGCCCTTCTTAATTCCTTGAGCCCTTCGACACTACTCTGAGAA 20055  
 Qy 2370 nGlyGluIleLeuProLeuLysThrLeuThrTrpValAlaLeuGlyValThrLeuAlaI 2390  
 Db 20056 TGGGAGATCTCGCCACTGAAGCACTGACATACGTGGCTTAGAGTGTCACTTGCTGCTG 20115  
 Qy 2390 aLeuLeuLeuThrPhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisG 2410  
 Db 20116 CTTTCTGCTCACTTCTTCTTCTCTCACTCTCTTGCTATCTGCGCTTCAACCAACCG 20175  
 Qy 2410 YIleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyI 2430  
 Db 20176 CATCCACGTAACTGACAGCTGCTCGGGGCTGAGCTGAGTGTCTTCTCTGGGAAAT 20235  
 Qy 2430 eAsnGlnAlaAspLeuPro----- 2436  
 Db 20236 CAACCAAGCTGACCTCTGTAAAGATGTCTTACGCCAGAAACGTGCCACTTCTTC 20295  
 Qy 2436 ----- 2436  
 Db 20296 AGGCGCCCTCCCAAGGCCCCCACTGGGACACCCCTGCTCTGACACATGAATCTAATAAG 20355  
 Qy 2436 ----- 2436  
 Db 20356 TGCCTAGTGACAGCCTGGCCCGAGGGTTCCTCTCTGTGTGCTCCCCGGGATCCCCAG 20415  
 Qy 2436 ----- 2436  
 Db 20416 CACCTGCTGTGGCCCAAGGCTTCCCTGAAGCAATTCCCAACCAAGGCCCTCTCCAT 20475  
 Qy 2436 ----- 2436  
 Db 20476 GCCTGACCCGAGCAGAGCCTGTGCTTGGGGGGGCCCCGGTGTGACCTGCGCC 20535  
 Qy 2437 -----PheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysT 2453  
 Db 20536 TGGGCCCTGAGTTGCTGTCAACAGTATGCAATCTGTGCACTTCTGTATCCTGTGCA 20595  
 Qy 2453 hPheSerTrpAlaLeuLeuGluAlaLeuHisIleLeuTyrArgAlaLeuThrGluValArgA 2473  
 Db 20596 CTTTCTGTGGGTCTGTGTGAGGAGCCCTTGACCTTGACCGGACCTCACTGAAGTCCGG 20655  
 Qy 2473 sPValAsnThrGlyProMetArgPheTyrTyrMetLeuGlyTrpGlyValProAlaPheI 2493  
 Db 20656 ATCTCAACAACCGGCCCAATGCGCTTCTATCAATGCTGGGCTGGGCGTGGCTTCA 20715  
 Qy 2493 IeThr----- 2494  
 Db 20716 TCAAGAGTATCCCAACCAATCCAGTCTTGGGGTCCCAATCCCTGGGTCCACTTTGT 20775  
 Qy 2494 ----- 2494  
 Db 20776 GCCATGTCTCTCAACCAATACAGGCGCTGAGGCCCAATCCCAATGCCCAAGCGG 20835  
 Qy 2494 ----- 2494  
 Db 20836 CTTTATTCACAGGTGTCCCTGTGTAAACCAAGCTGTGACGCGGCCCAACGAGCGTCA 20895





Db	23053	AGGAGAAAGAGAGAGCGCTTCCCTGGAAGACAGGAGCTGGATTACCTGTGGGCGCTGG	23111
Qy	2768	lylaaglwaqluaproleuHisSerThProlys-----	2779
Db	23113	GAGCAGAGAGACTGCGCTCTGACAGTACTCCAAAGGGTGGGCCAGCATGGGCGCTGTGGC	23172
Qy	2779	-----	2779
Db	23173	CTTTGGGGCCAGTGGAGGAGACATGTGGGCGCTGGGGTTCTTGGAGAGAGACTGGGGTGGT	23232
Qy	2780	-----	2782
Db	23233	GGCTCTCGCTATCTGCGCTCTGTGGGCGCTCATCTTCTTTCCCAACAGAGTGGGG	23292
Qy	2782	lyProgllyProgllyLyValAProTrProgllyAapPhegllyThThrAlaLySgluSers	2802
Db	23293	GCCCAAGGCGCTCGAGAGAGCGCCCTCGGCGAGAGACTTTGGAGACACAGCAAGAGAGATTA	23352
Qy	2802	erGlYaaNglyAlaProgluGluaProglleuArgluluaNglyAapAlaLeuSeraArglulug	2822
Db	23353	GTGGCAACGGGGCGCCCTGAGAGAGCGGTGGCGGAGAAATGAGATGCCCTGTCTCGAGAGG	23412
Qy	2822	lySerleuGlYProleuProgllySeraZlaGlnProHlslys-----	2836
Db	23413	GGTCCCTAAGCGCCCTTTCACAGGCTCTTCTGCCACGCTCAAAAGGTAGTGGGCAACC	23472
Qy	2836	-----	2836
Db	23473	CCAGCTGCGCAGACTCCCTATGTACAGACGCTCATACTCATTTCTCTGTGGCGGCAACC	23532
Qy	2837	-----	2849
Db	23533	TCACAGGCCCGGCGCCGAGCCCAAGGCAATCTTTAAGAAAGTGTCTGGCCACCATCAGAGG	23592
Qy	2849	lylySeraSerleuLeuArgleuProleuGlulugInCyvThrglySeraArglySers	2869
Db	23593	AGAAAGACAGCTCTCTGTGGGCTCCCCCTGAGACATGACAGAGGTCTTCCCGGCGCTCTCT	23653
Qy	2869	erAlaSerGlYlySeraArglyGlyProProProProArgProProProArgInSerleuG	2889
Db	23653	CCGCTAAGTGAAGGCGACGCGGGAGAGCCCGCTCCCGCCACGCGCCCGGAGAGGCTCCG	23712
Qy	2889	IngluGlnleuAaNglyValMetProIleAlaMetSerIleLyValaGlyThValAspG	2909
Db	23713	AGGAGACAGCTAAAGGGTCACTGCCATCGCATGACATCAAGGACAGGACGCTGGATG	23772
Qy	2909	lyAapSeraSerGlySerglyu 2915	
Db	23773	AGGACTCGTCAAGGCTCCGAG 23792	
RESULT 5			
US-09-737-149-1			
; Sequence 1, Application US/09737149			
; Patent No. US20020077466A1			
; GENERAL INFORMATION:			
; APPLICANT: Spaderna, Steven K			
; APPLICANT: Quinn, Kerry B.			
; APPLICANT: Shimketers, Richard A.			
; APPLICANT: Muralidhara, Padiganu			
; APPLICANT: Syplek, Kimberly A.			
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same			
; FILE REFERENCE: 15966-620 CIP			
; CURRENT APPLICATION NUMBER: US/09/737,149			
; CURRENT FILING DATE: 2001-06-15			
; PRIOR APPLICATION NUMBER: 60/170,564			
; PRIOR FILING DATE: 1999-12-14			
; PRIOR APPLICATION NUMBER: 60/173,165			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: 60/173,362			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: 60/173,544			
; PRIOR FILING DATE: 1999-12-29			
; PRIOR APPLICATION NUMBER: 60/174,404			

[illegible]

[illegible]

QY	518	ValProLeuNGlyTyrTLeuValLeuMetIstValGlnIalIeAspAlaAspAlaGlyIAspAsn	537
Db	1732	GTGGCCCTGGGCTACCCCGGTGGTGCATCTTCAAGCGCGTGGAGCGGAACTCTGGAAAGAAC	1799
QY	538	AlaArgLeuGluTyrArgLeuAla	549
Db	1792	GGCCGGCTGCACTATCGCTGGTGGACACGGCGTCCACTTCTTGGGGGGCGGACGGCT	1855
QY	550	-----AspPheProPheThrIleAsnAsnGlyThrGly	560
Db	1852	GGGCGCTAAGAAATCCGCGCCCAACCCCTGACTCTCCCTTCAGATCTCAACACAGCTCCGGT	1911
QY	561	TripIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal	580
Db	1912	TGGATCAACAGTGTGTGGCGAGCTGGACCGCGAGAGAGTGGAGACACTACGCTTCGGGGTG	1971
QY	581	GluAlaAspArgPheGlyThrProAlaLeuThrAlaSerAlaSerValThrVal	600
Db	1972	GAGGGGGTGGACCAACGGCTCGCCCCCAATGAGCTCTTCCACACAGCTGTCCATACCGGTG	2031
QY	601	LeuAspValAlaAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn	620
Db	2032	CTGAACGTGAATGACACACACCCCGGTGTTCACGACGGCCACTACAGACTTCGCTGAAAT	2091
QY	621	GluAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis	640
Db	2092	GAGGATGGCGCGGTGGGGAGCAGCGTGTGTGACCTCGACGGCGCCGACCGTGAAGCCCAAC	2151
QY	641	SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer	660
Db	2152	AGTGTGAATTACTTACACAGCTCAACGGCGGCAACCCGGAAACCGCTTGGCACTACAGACGC	2211
QY	661	GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrIleGluAlaGln	680
Db	2212	CAGAGAGGGGGCGGCGCTCATACCTGGCGTACTCTGGACTACACAGAGAACAGACAG	2271
QY	681	TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal	700
Db	2272	TACGTGTGGGGGTGAGACGATCCGAGCGGACACAGGTGCGACACTGGCGCAATGCTCTATTC	2331
QY	701	AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn	720
Db	2332	AACTTCATCTATGCGAACACCCACAGGCGCTGTCTTTCAGAGCTCCCACTTACACAGTAGT	2391
QY	721	ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp	740
Db	2392	GTCAAGTAGGACAGGGCTGTGGGCACTCTCAATGGCTCACTCCAGTGGCCACGATGAGGAC	2451
QY	741	ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle	760
Db	2452	ACAGAGAGAAATGCCCGCATCACTTAAGTGAATGAGACCCCGAGCGCAAGTTCGGCATT	2511
QY	761	AspAlaAspThrArgIalValIleThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer	780
Db	2512	GACCCCGCAGTGGCACCATGTATACACATGATGGAGCTGGACTATGAGAACAGAGTGGCC	2571
QY	781	TyrThrIleuAlaIleThrAlaArgAspAsnGlyIleProGlnIleYserAspThrThrTyr	800
Db	2572	TACACGTGTACCATCATGACGCGCAGGACAAAGGATCCCGAGAAATGACACACACACC	2631
QY	801	LeuGluIleLeuValAlaAspValAlaAspAspAlaProGlnPheLeuAspSerTyr	820
Db	2632	CTAAGAGATCTCATCCCTCGATGCCAAGACAAATGCAACAGCTTCTGTGGAGATTCTAC	2691
QY	821	GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr	840
Db	2692	CAGGGTTCATCTTTGAGATGTCTCCACCTTCACACGATCTCTCCAGGCTCTGGCACGG	2751
QY	841	AspArgAspSerGlyLeuAsnGlyValArgValPheTyrThrPheGlnGlyValAspAspGly	860
Db	2752	GACCGGGAATCAAGTCCCATGGGCGGTCTGTGTACCTTCCAGGGTGGGAGACAGCGGC	2811
QY	861	AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuAspArgLeuAsp	880

Db 2812 GATGGGAGCTTCTACATCGAGCCACGCTCGGTGATTTCGACACCGAGCGCGGTGAG 2871  
 Qy 881 ATGGUAAAGValAAGlnValValLeuArgAlaTyrAlaValAspArgIlePro---- 899  
 Db 2872 CGGGAAGATGGCCGTGATCAACCTTGGGCTCTGGCTGTGATGGAGGAGTCCACT 2931  
 Qy 900 ProAlaArgThrPrometGluValThrValThrValLeuAspValAspAspProPro 919  
 Db 2932 CCCCTTAGCGCCCTCGGAGAAATCCAGGTGACATCTTGACATTAATGACAAATGCCCC 2991  
 Qy 920 ValPheGluGlnAspGluPheAspValPheValGluGluAspSerProIleGlyLeuAla 939  
 Db 2992 ATGTTTGAGAAAGCAACTGAGCTGTTGTGTGAGAGAAACAACCAAGGGGTGTGTG 3051  
 Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3052 GTGGCAAAAGATTGTGCTTAACGACCTGTGTAAGGCCCTTAATCCCAAGTCAATATCAG 3111  
 Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3112 ATGTGGAAGAGGAGCAATGCGGCAATTCTTCACGTGACCTGCTCAACGGGAGCTGCGT 3171  
 Qy 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
 Db 3172 GGCATGTGAGAGCTTGACATTGAGGTCCGGCGGAGATATGTGTGTGTGTGTCGAGCGCACG 3231  
 Qy 1000 SerAlaProLeuValSerArgAlaThrValIleValArgLeuLeuAspArgAspAsn 1019  
 Db 3232 TCGGCTCCGCTGTGTGAGCCGAGCCAGGTGACATCTTCTGTGACCAAGATGACAA 3291  
 Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSer 1039  
 Db 3292 CCGCCCTGTGTGCGCCCACTTCCAGATCTCTTCAACAACATATGTCAACAAGTCAAC 3351  
 Qy 1040 SerPheProGlyGlyValIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 Db 3352 AGTTTCCCAACCGGCTGATCGGCTGTGATCCGAGCCATGACCCCAAGTGTGACAGC 3411  
 Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 Db 3412 CTCACTACACCTTTCGTCAGAGGCAAGAGCTCGCTGTGTGTGTGACCCCGGCAAG 3471  
 Qy 1080 GlyGluLeuLeuLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaIleMetSer 1099  
 Db 3472 GGGGAATGAGCTCAGCCCGGAGCTGGAACAACAACCGGCTGAGGGGCTCATGAGAG 3531  
 Qy 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
 Db 3532 GTGTCTGTCTGATGGCATCCACAGGTGACGGCTTCTGACCCCTGCGGTCAACATC 3591  
 Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139  
 Db 3592 ATACCGAGAGCAATGTCAGCAACAAGATCACTGCTCGCTGGAACAATGTCCAGAGAG 3651  
 Qy 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAlaThr 1159  
 Db 3652 AAGTTCTGTGCTCCGCTGTGCGCTTCTGTGAGAGGGGTGTGCGCGCTGTGTCTCACCC 3711  
 Qy 1160 ProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGlnHis 1179  
 Db 3712 ACCAAGAGCAAGCTTCTGCTTCACAGTCAAGAACACCAAGCTC--AGCTCCAAC 3768  
 Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhePro 1199  
 Db 3769 ATCTGAACAGTAACTTCTCGGCGCTGTGCTGTGCGGCTGCGGCTGCGGCTGCGGCT 3822  
 Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
 Db 3823 TTCCTCGGAGAGCACTGCGAGAGCAAGATCTTGAATGAGCGCTGTGACCAACATC 3882  
 Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239

Db 3883 TTCACGAGCGGCTGCTCCCTTCGACGACAACATCTGCTGCGAGCCCTGCGAGAAC 3942  
 Qy 1240 TyrMetArgCysValSerValLeuAspPheAspSerSerAlaProPheIleAlaSerSer 1259  
 Db 3943 TACATGAAGTGTGCTTCCGTTCTGCAATTCGAACTTCGAGCCCTTCTCTCACTCAAC 4002  
 Qy 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysAspProGly 1279  
 Db 4003 ACCGTGCTTCTCCGCGCCATCCACCCCAACAGGCTGCGGTGCGCTGCGCGCGGCG 4062  
 Qy 1280 PheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
 Db 4063 TTCACCGGAGACTCTGAGAGCGAGATCGAATCTCTCTCACTTCGACCCGTCGCGCGC 4122  
 Qy 1300 HisGlyArgCysArgSerArgGlyGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
 Db 4123 AACGCCCTCTGCGAGCCGAGGCGGCTACACTTCGAGTCTTGAGAGACTTCACT 4182  
 Qy 1320 GlyGluHisCysGlyValSerAlaArgSerGlyArgCysThrProGlyValCysValAsn 1339  
 Db 4183 GGAGAGCACTGTGAGGTGATGCTCCGCTCAGGCGCTGTGCAACGGGGGTGTGCAAGAAC 4242  
 Qy 1340 GlyGlyThrCysValAsnLeuLeuValGlyGlyPheLeuCysAspCysProSerGlyAsp 1359  
 Db 4243 GGGGCACTGTGTAACCTGTCAATCGAGCTTCCACTGTGTGTCTCTGTCGAG 4302  
 Qy 1360 PheGluLeuProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
 Db 4303 TATGAGAGCCCTTAATGTGAGGTGACACAGAGCTTCCCGCCCAATCTTCTGTCAAC 4362  
 Qy 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrLeuGlu 1399  
 Db 4363 TTCGGGGCGCTGAGACAGCGCTTCCACTTCACTCTCCCTCACTGCACTGCACTGAGAA 4422  
 Qy 1400 ArgAspGlyLeuLeuLeuThrAsnGlyValArgPheAsnGlyValHisAspPheValAlaLeu 1419  
 Db 4423 AGGAACGGCTGTCTCTTCAACCGCGCTTCAATGAAAGACAGACTTCACTGCGCTG 4482  
 Qy 1420 GlnValIleGlnGluGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
 Db 4483 GAGATCGAGAGCAAGAGAGTGAAGTCACTTCTGTGACGGGAGACAACAACAGACGTG 4542  
 Qy 1440 SerProPheValProGlyGlyValSerAspGlyGlnIleThrAlaThrValGlnLeuArgTyr 1459  
 Db 4543 GCACCGAAGGTCCAGTGTGTGATGAGTGAAGGCGGTGCACTGTGTGAGGTGAGTAC 4602  
 Qy 1460 TyrAsnLeuProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGluGlnVal 1479  
 Db 4603 TACACAAAGCCCAATATTTGGCCACTGGGCTGTGCCATGGGCTCCGGGGAAGAATG 4662  
 Qy 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 Db 4663 GCGGTGTGAGCAGTGAATGTGTGACAAACCAAGTGTGTGCGCTTGGAAAGAGATC 4722  
 Qy 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerValLeuSerLeuAspLeu 1519  
 Db 4723 GGAACCTACAGTGTGCGTCCGAGGCACTCAAGCGGCTCAAAAGTCCCTGTGATCTG 4782  
 Qy 1520 ThrGlyProLeuLeuLeuGlyGlyValProAspLeuProGluSerPheProValArgMet 1539  
 Db 4783 ACCGGCTCTTACTCTGCGGGGTGTCCCAACCTGCGCAAGAGATTCCTCACTGCAACAC 4842  
 Qy 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleLeuPheAla 1559  
 Db 4843 CGGCAAGTGTGTGAGTGAATGAGAACTGTGACGTGCAAGCAAAATGTGACATGTGCGC 4902  
 Qy 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaValAlaValAsnValCysAsp 1579  
 Db 4903 GGAATCATGCGCAAAATGAGCAACCGGGAAGCTGCGCTCGAGAGACTTCTGTGAGT 4962  
 Qy 1580 SerAsnThrCysHisAsnGlyGlyTyrCysValAsnGlnIleProAlaPheSerCysGlu 1599  
 Db 4963 GGAAGCGGCTGTCAAAATGAGAGCACTGTGTCAACAGTGAATATGTATCTGTGTAG 5022

QY 1600 CysProLeuGlyPheGlyGlySerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTCACTCCGATTCGGCGGAGAACTGTGACAAAGCCATGCTCAACCCCGACGCTTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTTPHIsGlyLeuSerLeuProHisSerGlnProTTPYr 1639  
 DB 5083 AGCGGTGAAGAGCGCTGTCTCGAGAGTGAACCTGAACATCATCTCTGTGCGCTGTGAC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 DB 5143 CTGGGGCTCATGTTCCGGACCCGGAAAGAGACAGCGTTCTTAATGAGAGCCACCAAGTGT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisIleValMetLeuSerValGlnGly 1679  
 DB 5203 GGGGCCACACAGCTTTCGCTCCAGATCTGAACATCACTCACTGTTGAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuAlaGlnLeuProGlyValArgAlaAsnArgGlyAspTTP 1699  
 DB 5263 GGGCCCTCCGATGTGAGTCCGTGATGTCTGCGGGTTCGGGGTGAACCAAGGGAGTGG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlnHisAlaIle 1716  
 DB 5323 CACCACTCTGTATCAAGCTGAAGAAATGTTAAGAGAGACATGACATGAAGCACCTGTGTC 5382  
 QY 1717 ---LeuSerPheAspTTPGlyGlnGlnArgAlaGlnGlyAsnLeuGlyProArgLeuHis 1735  
 DB 5383 ACCATGACCTTGATCAATAGGATGACACAAACAGACAGATATCGGGGGCATGCTTCCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlnGlyIleProGlyProAlaGlyValAla 1755  
 DB 5443 GGGCTGACCGTGAAGAGCGTGTGTGTGAGGGCGCTGAAGACAAAGTCTCCGTCGCC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAsn 1775  
 DB 5503 CGTGAATTCGAGGCTGATGACAGAGAGTGAAGATGGGGGGAGCGCCACCAAGCTCGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
 DB 5563 ACCCTGAACATGAACAACCACTCAAGTCAAGGTGAAGACGCGTGTATGTGACGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspThrAspSerTyr 1815  
 DB 5623 CCTGTACTCCAGACCCCTGTCCCAATAGCGCTGCCACACCCCTGGAGAGCATAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 DB 5683 AACTGGCTGTGTGACAAAGGTATCTTGATAATTAATGTGTGATGCCCTGTCACTGAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgGlyAspSerAlaProHisGlyTyrThrCys 1855  
 DB 5743 CCTGTGAACATGCGGGCGCTGTGCGCTCCCGGCTCCCGGAGGGTAACTGCTGC 5802  
 QY 1856 GlnCysProProAsnIleLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
 DB 5803 GAGTGTGGGCGCAAGTCAACGCGCGTACTGTGAACAAACTCAACTCTCCGTGCCCC 5862  
 QY 1876 ArgGlyTyrTTPGlyHisAspProThrCysGlyProCysAsnCysAspValSerIleGlyPhe 1895  
 DB 5863 AAGAGGTGTGGGGGAAACCCGCTGTGTGACCTGTGCACATGTGCGCTCAAGCAAGCTTT 5922  
 QY 1896 AspProAspCysAsnIleThrSerGlyGlnCysHisCysValGlnIleAsnIleTyrArgPro 1915  
 DB 5923 GATCCCGACGTAAATGAACCAACGCGCAAGTGCATGACAGAGAAATTAACAAGCTC 5982  
 QY 1916 ProGlySerProThrCysIleLeuLeuCysAspCysTyrProThrGlySerLeuSerVal 1935  
 DB 5983 CTAAGCCAGACACTGTGTCTGCGCTGTGACCTGTCCCAATGAGCTCCCAACGCGCACT 6042  
 QY 1936 CysAspProGlnAspGlyGlnCysProCysValProGlyValIleGlyArgGlnCysAsp 1955  
 DB 6043 TCGACATGCGCAACCGGGAGTGTGCTCTCAAGCCCGGGGTCAATGGCGCGCAAGTGCAC 6102

QY 1956 ArgCysAspAsnProPheAlaGlnValThrThrAsnGlyCysGlnValAsnTyrAspSer 1975  
 DB 6103 CCTGTGACAAACCGTTTGTCCAGGTCACACAGCTCTGGTGTGAATGATCTTCAATGCC 6162  
 QY 1976 CysProAlaAlaIleGlnAlaGlyIleTTPProAspThrArgPheGlyLeuProAla 1995  
 DB 6163 TGTCCCAAGCATTTTGAAGCCCGGCACTGTGTGCCACAGACCAAGTTCCGGGACGCGCT 6222  
 QY 1996 AlaAlaProCysProValSerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 DB 6223 GCGGTGCCATGCGCTTAAGGATTCCTTGAAATGCGGTGCACATGCACTGCACGCGGAGAAAG 6282  
 QY 2016 GlyTTPLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuGlyGly 2035  
 DB 6283 GGTGGCTGCGCCCGCAGAGCTTTTAAGCTGTACCAACATCTCTGTGTGACCTCAAGGGCC 6342  
 QY 2036 PheAlaGlnArgLeuGlnArgAsnGlyLeuAspSerGlyValArgSerGlnGlnLeu 2055  
 DB 6343 ATGAATGAAAGACTGAGCGCAATGACACGAGTGTGACGCGCGCAAGGCGCTGACGCTG 6402  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIys 2075  
 DB 6403 GTGAGGGCGCTGTGAGTGTACACACACACAGCGCAAGCTTGTGGCATATAGCTGCC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 DB 6463 ACGGCTTACAGCTGTGGGCGACAGCTCTTGACAGACAGAGCTGACAGAGGCTTGCAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlyLeuLeuLeuArgValGlySerAlaLeu 2115  
 DB 6523 CTGGCAGCGACGACGAGACGCGCATTTCAACAGACAGCATCTCACTCGGACGCGCTTC 6582  
 QY 2116 LeuAspThrAlaAsnIleValArgHisIleThrGlyLeuIleGlnGlnGlyGlyThrAla 2135  
 DB 6583 CTGGCGCCAGACCAACAGGCGCGCTGGAGACAGATCCAGCGAGAGAGGCGGACGCGCA 6642  
 QY 2136 TrpLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThr 2155  
 DB 6643 CAGCTGTCCGCGCGCTGAGGGCTAATTCAGACAAAGTGAAGCGCAAGTGTGGCGGACG 6702  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175  
 DB 6703 TACCTGCGGCGCTTGTGATGTATCGCCCAACATGATCTTGTGCTGACATCTTTGAC 6762  
 QY 2176 IysGlyAsnPheAlaGlyAlaIleLeuProArgTyrGlnAlaLeuArgGlyGlyGlnPro 2195  
 DB 6763 AAGTCAACTTTAAGGAGAGGAGGCGCGGATCCCGGATTCACACATCAAGAAAGTTCGCC 6822  
 QY 2196 ProAspLeuGlnThrThrValIleLeuProGlnSerValPheArg-----Gln 2211  
 DB 6823 AAGGAGCTGAGTCTCCGTCTCTCCAGCCGCACTTCTTCAAGACCACTGAAGAAAAA 6882  
 QY 2212 ThrProProValValArgProAlaGly-----ProGly 2222  
 DB 6883 GAGGCGCCCTGTGAGCGCGCTGTGCGCGGAGAGACACCCGCGACCAACGCGCGCGGGG 6942  
 QY 2223 GlnAlaGlnGlnProGlnGlnLeuAlaArgArgGlnArgHisAspProGlyLeuSerGln 2242  
 DB 6943 CCGGCAACGAGAGAGAGAGGCGCCCATACAGACGAGGAGGAGACACCTGTATGACGCTGCC 7002  
 QY 2243 GlyGlnAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsn 2262  
 DB 7003 CAGTTGCGCGCTGTGCTGTATCACTTAACCGCACCTGTGGGAGAGCTCTGCGCCAGGCGC 7062  
 QY 2263 TyrAspProAspValArgSerLeuArgValProIysAspProIleIleAsnThrProVal 2282  
 DB 7063 TAGACCCCGACCGTGTGAGCTGTGCTGTGCTCAACGCGCATATTAATACCCGATG 7122  
 QY 2283 ValSerIleSerValHisAspAspGlnGlnLeuLeuLeuProArgAlaLeuAspValProVal 2302  
 DB 7123 GTGACAGCGGTGTATACAGAGAGGCGCTCGGCTCCGAGACCCCTGGAAGAGCGCGTTC 7182  
 QY 2303 ThrValGlnPheArgLeuLeuGlnThrGlnIleArgThrIlyAspProIleCysValPheTTP 2322

Db	7183	CTGGGAGAGTTTCGACCTCTGGAGAGTGGAGGAGGACCAAGACCTGTGCGTGTCTCG	7242
Qy	2323	Aenh1sser11eleuValSerGly1YThr1yGly1TTPSer11ahArgGly1CyGluValVal	2342
Db	7243	AACCACTCCCGGCGGTGGTGGAGCGGAGAGGTGTGCGCGGAGCTGGAGACTCTCG	7302
Qy	2343	PheArgAnGluSerNH1ValSer1CyGln1CyAsnNH1smetThrSerPheAlaValLeu	2362
Db	7303	TCCAGGAACCGGACACATGTGCGCTGCCAGTGGACGCAACAGCACGCTTTGGCGTGTCTC	7362
Qy	2363	MetArgValSer1hArg1Glu1Asn1yGlu1Ileu1Pro1Leu1yThr1Leu1Thr1yVal	2382
Db	7363	ATGGATATCTCAAGGCGTGAAGACGGGAGGCTCGCTCTGAGATGTGCATTATGCC	7422
Qy	2383	AlaLeuGlyVal1Thr1Leu1Ala1Leu1Leu1Thr1PhePhePheLeu1Thr1Leu1Arg	2402
Db	7423	GCTGTGCTCTGTCACTGGACGCTGTGCTGGCGCTTTCGTCCTTGAAGCTGTGCTGC	7482
Qy	2403	I1eleuArgSerAnGlnH1sgly1LeaArg1ArgAnLeu1Thr1Ala1LeuGly1Leu1Ala	2422
Db	7483	ATGCGACGCTCCAACTGCACAGCAATTCAACAGACCTCGCGAGGCGCTCTCTCTCTCT	7542
Qy	2423	Gln1Leu1Val1Phe1Leu1Leu1y1I1eAsnGln1Ala1sPleu1Pro1Phe1Ala1Cy1Thr1Val1Ile	2442
Db	7543	CAGCTGGTGTTCGATATGGAGATCAACACAGAGAAACCCGTTTCTGTGTGCAGATGGTT	7602
Qy	2443	Ala1Ile1Leu1LeuNH1sPhe1Leu1y1Leu1Cy1Thr1Phe1Ser1TPAla1Leu1Glu1Ala1Leu	2462
Db	7603	GCCATCTCTCTCCCACTACATCTACATGAGACCTTTGCTGTGAACCTTCGTGTGAAGACTGTG	7662
Qy	2463	Hi1Leu1y1Arg1Ala1Leu1Thr1Glu1Val1Arg1Asp1Val1Asn1Thr1y1Pro1Met1Arg1Phe1y1r	2482
Db	7663	CATGTCTCAACCGCAATGCTGACCGAGGTGGCAACATCGACAGCGGAGCCCATCGGTTCTTAC	7722
Qy	2483	Ty1Met1Leu1y1TPG1y1Val1Pro1Ala1Phe1I1e1Thr1Gly1Leu1Ala1Val1Gly1Leu1Asp1Pro	2502
Db	7723	TACCTCGTGGGCTGGGGCATCTCCGGCATTTGTCAAGACCTGGCGGTGCGCTGTGAACCC	7782
Qy	2503	Glu1Gly1y1y1Arg1An1Pro1Asp1Phe1Cy1Thr1Leu1Ser1I1e1y1Asp1Thr1Leu1I1e1TP1Ser	2522
Db	7783	CAGGCGTACCGGAACCCCGCACTTGTGCTGCGTGTGCTTCAAGACACCTCATATTGGAGCG	7842
Qy	2523	Phe1Arg1y1Pro1Val1Ala1Phe1Ala1Val1Ser1Met1Ser1Val1Phe1Leu1y1r1I1e1Leu1Ala1Ala	2542
Db	7843	TTTGGCGGAGCCCATCGAGACTGTTATATATCAACACAGTCACTTGTCTCATATGTGCA	7902
Qy	2543	Arg1Ala1Ser1Cy1Val1Ala1Gln1Arg1Gln1y1Phe1Glu1y1y1y1Gly1Pro1Val1Ser1Gly1Leu	2562
Db	7903	AAGGTTTCTCTGCCAAAGAAAGCACACATTATATGGAAAAAAGGAGTGTCTCTCTGTG	7962
Qy	2563	Gln1Pro1Ser1Phe1Ala1Val1Leu1Leu1Leu1Ser1Ala1Thr1P1Leu1Leu1Ala1Leu1Leu1Ser	2582
Db	7963	AGGACCGGATTTCTCTGTGCTGTCTGTCAATCAAGCGCACCTGGCGTGGGGCTGTGCTGTGCTG	8022
Qy	2583	Val1Asn1Ser1Arg1Thr1Leu1PheNH1s1y1y1y1Leu1Phe1Ala1Thr1Cy1Asn1Cy1I1e1Gln1y1	2602
Db	8023	GTTAAACCGGAGTGAAGCTTACAGCTTTCACATCTTTCGCACTTCACGCGCTTACAGGCG	8082
Qy	2603	Pro1Phe1I1e1Phe1Leu1Ser1y1r1Val1Val1Leu1Ser1y1Glu1Val1Arg1y1Ala1Leu1y1--	2621
Db	8083	CCCTTTCGCTCTCTTTTCCACTGGGTGTCTCAACAGAGGAGTCCGGAAGGACCTGAAGGCGC	8142
Qy	2622	Leu1Ala1Cy1Ser1Arg1y1r1Pro1Ser1Pro1Arg1Pro1Ala1Leu1Thr1Thr1y1Ser1Thr1---Ileu	2640
Db	8143	GTCGTCGGCGGAGAGAACTGCACTGGAGGACTTCGCGACACCAAGGCACTCTGTGCTG	8202
Qy	2641	Thr1Ser1Ser1y1r1y1Asn1Cy1Pro1Ser1Pro1y1r1Ala1Arg1y1-----Arg1Leu1y1r1Gln1Pro	2658
Db	8203	ACGCGACTCCCTCAACTGCACACACACTTCGATGACGGGCTGTGACATGTCTGCGACAGAC	8262
Qy	2659	Ty1r1y1Arg1Ser1r1Ala1y1Ser1LeuNH1Ser1Thr1Ser1Arg1Ser-----	2672

[illegible]







```

Db      3769 ATCTGAAAGTAACTTCTCGGCGCTGCGCTCGGCGGCGTCCGCGC-----CAGTTTC      3822
Qy      1200 LeuprSerSgluAapLeuGlnGlnAargLeuYrLeuAAsnAargSerLeuThraIle      1219
Db      3823 TTCCTCGGAGAGACCTGCAAGAGACAGATCTACTAAATCGAACCTGCTTACACCACT      3882
Qy      1220 SerAlaGlnAryValLeuProPheAapAAsnIleCyAseuAryGluProCySgluAan      1239
Db      3883 TCCACCCACCGCGCTGCTCCCTTGCACGACCAACATCTGCTGGCGAGCCCTGCGAGAAC      3942
Qy      1240 TyrMetAryCyValSerValLeuAargPheAapSerSerAlaProPheIleAlaSerSer      1259
Db      3943 TACATAAAGTGGGTGCTGCTTCTGCGATTGCAAGCTGCGCGCTTCTCTCAGCTCCAC      4002
Qy      1260 SerValLeuPheAapProIleHisProValGlyGlyLeuAryCyAargCySapProGly      1279
Db      4003 ACCGTGCTTTCCTCGGCGCATCCACCCCAACAGCGCTGCGCTGCGCTGCGCGCGC      4062
Qy      1280 PheThrGlyAapTyrCySgluThrGlnValAapLeuCyArySerAargProCySglPro      1299
Db      4063 TTCACCGCGGACTACTGCGAGAGAGATGACCTGCTACTCCGACCCGTCGCGCGC      4122
Qy      1300 HisGlyAargCyAargSerAryGlyGlyTyrThrCyAseuCyAargAapGlyTyrThr      1319
Db      4123 AACGCGCGCTGCGCGAGCGCGAGCGGCGCTACCTCGAGAGTGTTCAGAGACTTCACT      4182
Qy      1320 GlyGlnHisCyAseuGlnValSerAlaAargSerGlyAargCyAthrProGlyValCyAlyAAsn      1339
Db      4183 GGAAGAGACTGTAGAGTGAATGCCCTCAAGCGCTGAGCCCAACGAGGATGACAGAAC      4242
Qy      1340 GlyGlyThrCyAValAAsnLeuValGlyGlyPheAryCyAapCyAProSerGlyAap      1359
Db      4243 GGGGCGCACTGCGGTAACTGCTCACTCGCGGCTTCACTGCGTGTCTCTCGCGAG      4302
Qy      1360 PheGlnAryProTyrCyAseuValThrThrAargSerPheProAlaHisSerPheIleThr      1379
Db      4303 TATGAGAGGCGCTACTGTAGAGTGAACACAGAGAGTTCGCCGCCAGTCTTCGTACCC      4362
Qy      1380 PheAryGlyLeuAryGlnAargPheHisPheThrIleAlaLeuSerPheAlaThrIleGlu      1399
Db      4363 TTCGCGGCGCTGAGACAGCGCTTCACTTCACTCACTCTCTGCGAGCGAGAACACACCGTG      4422
Qy      1400 AArgAapGlyLeuLeuLeuTyrAangGlyAargPheAangGlyAAsnHisAapPheValAlaLeu      1419
Db      4423 AGGAACGGCTGCTCTCTTCAACACGGCGCTTCAATGAGAACACACACTTCACTCCCTG      4482
Qy      1420 GlnValIleGlnGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrVal      1439
Db      4483 GAGATCGTGAAGAGACGAGGTGAGCTCACCTTCTGCGAGCGAGAACACACACGACCGTG      4542
Qy      1440 SerProPheValProGlyValSerAapGlyGlnThrPheIleThrValGlnLeuLeuTyr      1459
Db      4543 GCAACGGAAGTTCCTCAAGTGTGTAGTGAACGAGCGGTGCACTGTGACAGGTGCACTAC      4602
Qy      1460 TyrAAsnAryProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnVal      1479
Db      4603 TACAACAAGCCCAATATATGGCCACCTGGGCTGCCCAATGAGCGCGTCCGAGAAAGATG      4662
Qy      1480 AlaValValThrValAapGlyCyAapThrGlyValAlaLeuAargPheGlySerValLeu      1499
Db      4663 GCGGTGTGAAGATGATATGTGACACACATGCGTGTGCTTGTGAAAGACATC      4722
Qy      1500 GlyAAsnTyrSerCyAlaAlaGlnGlyThrGlnGlyGlySerGlyArySerLeuAapLeu      1519
Db      4723 GGAAGCTACAGCTGAGCTGCCAGAGGCACTCAGACCGGCTCAAAAGTCCCTGATCTG      4782
Qy      1520 ThrGlyProLeuLeuGlyGlyValProAapLeuProGlySerPheProValAargMet      1539
Db      4783 ACCGCGCTCTACTCTCTGCGGGGTGTCCCAACCTGCGCAGAAAGACTTCCCAATGCAAC      4842
Qy      1540 AArgGlnPheValGlyCyAAsnAryAAsnLeuGlnValAapSerAryHisIleAapPheAla      1559

```

---

```

Db      4843 CGGAGTTCTGGGCTGCATGCGGAACCTGTCACTGCAAGCGCAAAATGTGACATGCGC      4902
Qy      1560 AapPheIleAlaAAsnAangGlyThrValProGlyCySProAlaAlyAAsnValCyAap      1579
Db      4903 GGAATTCATCGCCAAATAGGACCCCGGAAGGCTGCGCTCGGAGGACTTCTGCGAT      4962
Qy      1580 SerAAsnThrCyHisAAsnGlyGlyThrCyValAangGlnThrAapAlaAAsnSerCySglu      1599
Db      4963 GGAAGCGGTGTCAAGATGAGAGGCACTGTGTCAACAGGTGAATATGATCTGTGTAG      5022
Qy      1600 CySProLeuGlyPheGlyGlyArySerCyAlaGlnGlnMetAlaAAsnProGlnHisPhe      1619
Db      5023 TGTTCATCTCCAAATCTGGGGAAGAACTGTGAGCAAGCAATGCTCCACCCCACTTTC      5082
Qy      1620 LeuGlySerSerLeuValAlaThrHisGlyLeuSerLeuProIleSerGlnProThrTyr      1639
Db      5083 AGCGTGAAGAGCTGTCTCTGAGTGAACCTGAACATCAATCATCTGTGCTCTGTGAC      5142
Qy      1640 LeuSerLeuMetPheAryThrAryGlnAlaAapGlyValLeuLeuGlnAlaIleThrAry      1659
Db      5143 CTGGGGCTCATGTTCCGGAACCGGAAGAGAGCAAGCTTGTGATGAGAGCCACAGTGT      5202
Qy      1660 GlyAargSerThrIleThrLeuGlnLeuAryGlnGlyHisValIleMetLeuSerValGly      1679
Db      5203 GGGCCCAACACTTTCGCTCGCATGCTGAACCACTACTCCAGTTGAGGTGTCCAC      5262
Qy      1680 ThrGlyLeuGlnAlaSerSerLeuAryGlnGlyProGlyAryAlaAAsnAapGlyAapThr      1699
Db      5263 GCGCCCTCCAGTGTGAAGTCCGTATGCTGTCCGGGTTCGGGTGACCGAGGAGTGG      5322
Qy      1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle      1716
Db      5323 CACCACTGTGATGCAAGCTGAAGATGTAAAGAGAGCAAGTGAATGAACACTGTGTC      5382
Qy      1717 --LeuSerPheAapTyrGlyGlnGlnAargAlaGlyAAsnLeuGlyProAargLeuHis      1735
Db      5383 ACCATGACCTTGTGACTATGGAGTGAACCAAGGCAATATCGGGGAGATGCTTCC      5442
Qy      1736 GlyLeuHisLeuSerAAsnIleThrValGlyGlyIleProGlyProAlaGlyValAlaIle      1755
Db      5443 GGGCTGACGGAAGAGCGGTGTGTGAGGCGCTCTGAAGACAGAGTCTCGTGGC      5502
Qy      1756 AArgGlyPheAryGlyCyAseuGlnGlyValAargValSerAapThrProGlnGlyValAAsn      1775
Db      5503 CGTGATTTCCAGGCTGTGATGACAGAGTGAAGATGGGGGAGGACGCCACCAACGTCCGC      5562
Qy      1776 SerLeuAapProSerHisGlyGlySerIleAAsnValGlnGlnGlyCySserLeuProAap      1795
Db      5563 ACCGTGAACATGAACACGCACTCAAGGTCAAGGTGAAGACGCGCTGTGATGTGACGAC      5622
Qy      1796 ProCyAapSerAapProCySProAlaAAsnSerTyrCySserAAsnAapThrAapSerTyr      1815
Db      5623 CCTGTACTTGAGACCTCTGTCTCCCAATAGCCGCTGCAACAGCGCTGGAGAGACTAC      5682
Qy      1816 SerCySserCyAapProGlyTyrTyrGlyAAsnAryCyThrAAsnValCyAapLeuAAsn      1835
Db      5683 AGTGTGCTGTGACAAAGGTGACTTGAATAACTGTGTGAGATGCTGTGACCTGAC      5742
Qy      1836 ProCySgluHisGlnSerValCySThrAryAryProSerAlaProHisGlyTyrThrCyS      1855
Db      5743 CCTGTGAAGAACAGGGGCTGTGCGCTCCCGGCTCCCGGAGGGCTGATGTGTGTC      5802
Qy      1856 GlnCySProProAAsnTyrLeuGlyProTyrCySgluThrAryGlyLeuAapGlnProCySPro      1875
Db      5803 GAGTGTGGGCCCACTCACTACGCGGCGGTATCTGTGAAGAAACATCTGACCTTCCGTGCC      5862
Qy      1876 AArgGlyTyrThrGlyHisProThrThrCyGlyProCyAAsnCyAapValSerGlyGlyPhe      1895
Db      5863 AGAGGCTGTGGGGAACCCGCTGTGTGACCTTGCACATGTGTGCGCTGACGAAAGGCTTT      5922
Qy      1896 AapProAapCyAAsnAryThrSerGlyGlyCyHisAryCyAlyAAsnHisAryAryPro      1915
Db      5923 GATCCCACTGTATTAAGACCAACGAGGCAAGTGCACATGTCAAGGAATTACTTACACAGCTC      5982

```

QY 1916 ProGlySerProThrCysIleuLeuCysAspCysTyr-ProThrGlySerLeuSerArgVal 1935  
DB 5983 CTAGCCAGGAGACCTGTCTGCGCCCTGCGAGCTGCTCCCACTGCTCCCACTGCGGACCT 6042  
QY 1936 CysAspProGluAspGlyGlnCysProCysValProGlyVal11IleGlyArgGlnCysAsp 1955  
DB 6043 TGGGACATGGCCACCGGCGAGTGTGCTCGACAGCCCGGCGGTCACTCGCCCGCTGACAC 6102  
QY 1956 ArgCysAspAspProPheAlaGluVal1ThrThrAsnGlyCysGluValAsnTyrAspSer 1975  
DB 6103 CGCTGGACACACCGCTTGGCGAGTCAACACCGCTGCGGTGAAGTGAATCTCAATGGC 6162  
QY 1976 CysProArgAla1IleGluAlaGly11ETPTTPProArgThrArgPheGlyLeuProAla 1995  
DB 6163 TGTCCCAAGACATTTAGAGCGCGCATCTGTGGCCACAGACCAAGTTCCGGGACGGGCT 6222  
QY 1996 AlaAlaProCysProGlySerPheGlyThrAlaValArgHisCysAspGlu1HisArg 2015  
DB 6223 GCGGTGCGACCTGCTAAGGATCGGTGMAATGCGGTCCGACATGCGAGCGGGAGAG 6282  
QY 2016 GlyTTPLeuProProAsnLeuPheAsnCysThrSer1IleThrPheSerGlyLeuGly 2035  
DB 6283 GGTGTGCTGCGCCACAGAGCTTTAACTGTACCAACATCTCTTCTGTGACCTCAAGGCC 6342  
QY 2036 PheAlaGluArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnLeu 2055  
DB 6343 ATGAATAGAAAGCTGAGCGCGCAATGAGACGCAAGTGAACGCGCGCGCTGCGAGCTG 6402  
QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGly1TyrPheGlySerAspVal1 2075  
DB 6403 GTGAGGCGCTGCGAGGTCTACACGACACCGGCGCGCTCTTGTGCAATGACGTGCC 6462  
QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
DB 6463 ACGGCTACAGCTGCTGCGGCGCACGTCCTTCAACAGAGAGCTGGAGAGCGCTTGCAC 6522  
QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlnAsnLeuLeuArgVal1GlySerAlaLeu 2115  
DB 6523 CTGGCAGCCACGAGACCGCGCATTTCAACGAGAGAGTCACTGCGGAGCGCGCTC 6582  
QY 2116 LeuAspThrAlaAsnLysArgHisTTPGlyLeu1IleGlnThrGlnGlyGlyThrAla 2135  
DB 6583 CTGGCCCGACGACACGAGCGCGCTGGAGACATTCAGCGAGCGAGCGGACGAGCA 6642  
QY 2136 TTPLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThr 2155  
DB 6643 CAGCTGCTCGGCGCTCGAGGGCTACTTCAGCAAGTGGACGCAACGCGGCGAGAG 6702  
QY 2156 TyrLeuSerProPheThr1IleVal1ThrProAsn1IleVal1IleSerValValArgLeuAsp 2175  
DB 6703 TACCTCGGCGCTTCGTCACTGTCACCGCCACATATCTTCTGTCCATCTTTCAG 6762  
QY 2176 LysGlyAsnPheAlaGlyAlaLysLeuProArgTyrGlnAlaLeuArgGlyGlnPro 2195  
DB 6763 AAGTTCACTTACGGAGACGAGGTCGCGCATTTGCAACATCCATGAGAGTTCC 6822  
QY 2196 ProAspLeuGlnThrThrVal1IleLeuProGlySerValPheArg-----Gln 2211  
DB 6823 AGGAGACGAGAGTCCGCTCTCTCCAGCGCATTTCTTCAAGCACTGAGAGAAAA 6882  
QY 2212 ThrProValValArgProAlaGly-----ProGly 2222  
DB 6883 GAAAGCCCTGCTGAGGCGCGCTGGCGGAGACCAACCCGAGACCAACGCGCGGGG 6942  
QY 2223 GluAlaGlnGluProGluGluLeuAlaArgArgGlnArgArgHisProGluLeuSerGln 2242  
DB 6943 CCGGACGAGAGAGGAGCGCCCGCATCAGAGCGGAGAGACCTGTATGACGTGGC 7002  
QY 2243 GlyGluAlaValAlaSerVal1IleTyrArgThrLeuAlaGlyLeuLeuProHisAsn 2262  
DB 7003 CAGTTGCGCGTCCGCTGTGTCACTTACCGCACCTGCGGAGAGCTCTGCGCGAGCC 7062

QY 2263 TyrAspProAspLysArgSerLeuArgValProLysArgPro1IleLeuAsnThrProVal 2282  
DB 7063 TACGACCCCGACCGGTGCGAGCTCCGGTTCCTTCAACCGGCGCATATTAATACCGCATG 7122  
QY 2283 ValSer1IleSerValHisAspAspGluGluLeuLeuProArgAlaLeuAspLysProVal 2302  
DB 7123 GTAGACACGCTGTATACGAGGAGGCGCTCCGCTCCGAGACCCCTGAGAGCGCGCTC 7182  
QY 2303 ThrValGlnPheArgLeuLeuGlnThrGlnGluArgThrLysPro1IleCysValPheTTP 2322  
DB 7183 CCGGTGAGAGTTGCGCTGCGAGGTGAGAGCGACCAACAGCTGTCTGCGGTTCG 7242  
QY 2323 AsnHisSer1IleLeuVal1SerGlyThrGlyTTPSerAlaArgGlyCysGluValVal 2342  
DB 7243 AACCACTCCCTGCGCGGTGTGTGGAGCGGAGGAGGTGTGCTCCGCGGCGCGAGCTCTG 7302  
QY 2343 PheArgAsnGlySerHisVal1SerCysGlnCysAsnHisMetThrSerPheAlaValLeu 2362  
DB 7303 TCCAGAACCGGACACATGTGCTGCGAGTGCACGCCACACAGCAAGCTTTCGCGTCTC 7362  
QY 2363 MetAspVal1SerArgArgGluAsnGlyGlu1IleLeuProLeuLysThrLeuThrTyrVal 2382  
DB 7363 ATGATATCTCCAGGCGGTGAGAGCGGAGAGTCTGCTCTGTAAGATGTCACTATGCC 7422  
QY 2383 AlaLeuGlyVal1ThrLeuAla1AlaLeuLeuThrPhePhePheLeuThrLeuLeuArg 2402  
DB 7423 GCTGTGCTCTGTCACTGAGCGCTGCGTGGGCTTCTGCTCTGAGCTGTGCTGCC 7482  
QY 2403 IleLeuArgSerAsnGlnHisGly11LeArgArgAsnLeuThrAla1AlaGlyLeuAla 2422  
DB 7483 ATGCTGCGCTTCCACCTGACGACATTCACAGACCTGCGCGGTGCGTCTTCTCTCT 7542  
QY 2423 GlnLeuVal1PheLeuLeuGly11LeuAsnGln1AspLeuProPheAlaCysThrVal1Ile 2442  
DB 7543 CAGCGGTGTTGTGATTTGGATTCACACAGACGAAACCGCTTCTGTGACAGATGCTT 7602  
QY 2443 Ala1IleLeuLeuHisPheLeuTyrLeuCysThrPheSerTTPAlaLeuLeuGluAlaLeu 2462  
DB 7603 GCCATCTCTCTCCATCACTATCACTAGAGACCTTGTGCTGACCTCGTGAAGAGCTG 7662  
QY 2463 HisLeuTyrArgAla1AlaLeuThrGluValArgAspValAsnThrGlyProMetArgPheTyr 2482  
DB 7663 CATGTCTACCGATCTACCGAGGTGCGCAACATTCAGACCGGCGCATGTGGTCTTAC 7722  
QY 2483 TyrMetLeuGlyTTPGlyValProAlaPhe1IleThrGlyLeuAlaVal1GlyLeuAspPro 2502  
DB 7723 TACGTGAGGCTGAGGATCCCGGCATTTGTCAGAGACTGGCGGTGAGACCC 7782  
QY 2503 GlnGlyTyrGlyAsnProAspPheCysTTPLeuSer1IleTyrAspThrLeu1IleTTPSer 2522  
DB 7783 CAGGCTACGGGAGACCCGACCTTGTGCTGCTGTGCTCAAGACACCTGTGATTTGAGGC 7842  
QY 2523 PheAlaGlyProValAla1PheAlaVal1SerMetSerVal1PheLeuTyr1IleLeuAla1 2542  
DB 7843 TTTGGCGGCGCATATGAGCTTTATATCATCAACAGTCACTTGTCTATCTGCA 7902  
QY 2543 ArgAlaSerCysAla1AlaGlnArgGln1PheGlyLysGly1ProVal1SerGlyLeu 2562  
DB 7903 AAGTTTCTGCGCAAGAAAGACACATATATGAGAAAGAGGATGTCTCTCTGCTG 7962  
QY 2563 GlnProSerPheAlaVal1LeuLeuLeuLeuSer1AlaTTPLeuLeuAlaLeuLeuSer 2582  
DB 7963 AGGACCGCATTTCTCTGCTGCTCACTACGCGCACCTGCTGCGGTGCTGCTGCT 8022  
QY 2583 ValAsnSerAspThrLeuLeuPheHisTyrLeuPheAla1ArgHisCys1IleGlnGly 2602  
DB 8023 GTGAACCGGATGACATGAGCTTACATACCTTCTGCGCATCTTCAAGGGCTTACAGGGC 8082  
QY 2603 ProPhe1IlePheLeuSerTyrValVal1LeuSerLysGluValArgLysAlaLeuLys--- 2621  
DB 8083 CCTTGTCTCTCTTTCACCTGCTGCTCAACAGAGGAGTCCGAGAACACCTGAGAGGC 8142  
QY 2622 LeuAlaCysSerArgLysPheSerProAspProAlaLeuThrThrLysSerThr---Leu 2640



QY 71 -----ArgCysArgAspAla-glyThrGluLeuT 80  
 Db 229 CGGCGCGTCTCGGCGCGGCGCGCGCTGCGCGTGCAGATCGCGTTGGT----- 278  
 QY 80 hrcgLYHleuValProHisHleuArgValThrCysProGluSerGluAlaH 100  
 Db 279 ---GCGCCGAGAGCGCC-----GAGCGGCTGAGCGG-----CGCGTGGCGGCGCC 324  
 QY 100 hIleProLeuProProAlaProGlu-glyCysProTrpSerCysArgLeuGluTyrIle 119  
 Db 325 AC-----GCACCTTCCCGGCTGCGGAGCGCGGCTGCGCGGAGAAC 369  
 QY 120 GYGLYHleuSerProGlnGlyYleuThrLeuPro----- 132  
 Db 370 GGTGCGCGGCTCTGC-----GGGCGCTCTGCTTCCCGTCCGCGCGGCTGCGCGGCG 423  
 QY 133 GlnGluHleuProCysLeuYleuVala----- 140  
 Db 424 GCGGAGCATTCGCGCGCTCGCAGCTCCGAGCACCTTACCGCGCTGCGCGCGCGCGC 483  
 QY 141 ProArgLeuArgCysGlnSer----- 147  
 Db 484 CCGAGGCGCGCGCTGCTCCGCGCGTCCCATCTGCTGCGCGCGGCGGCTCGGTCGCGCTG 543  
 QY 148 -----CysLYleuAlaGlnAlaProGly---LeuArgAlaGly----- 159  
 Db 544 CGTCTGCTGTGCGCGCTGCGCGCGCGCGCTGCGCGCGCTGCGCGCTGCGAG 603  
 QY 160 -----GluArg 161  
 Db 604 GCGCGCACCGCGCGGAGCGCGCTCGCGTCCGATCCCATCCGCGCGCTGCGCGCGAAC 663  
 QY 162 SerProGlnGluSerLeuGlyArgArgGlySerArgAsnValAsnThrAlaPro----- 179  
 Db 664 TTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGG 723  
 QY 180 -----GlnPheGlnProProSerTyrGlnAlaThrValProGlnAsnGlnProAlaGly 197  
 Db 724 AGCGTGAAGTTCCTCCAGTGCACATACAGGTGGCTTTGTGAGAACCAACCGCGGCGC 783  
 QY 198 ThrProValAlaSerLeuArgAlaIleAspProAspGlnGlyGluAlaGlyArgLeuGlu 217  
 Db 784 ACCCTCATCTCCAGCTGACGCGCGCATACCATTCGAGGCGAGGAGAGAGCGCGCTGAGC 843  
 QY 218 TyrThrMetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProVal 237  
 Db 844 TATTTACATGAGGCGCTGTTTCCAGCGAGCGCTCCCGGCGTACTTCCGAATTCACCTGCGC 903  
 QY 238 ThrGlyAlaValThrThrAlaGluGluLeuAspArgGluThrTyrSerThrHisValPhe 257  
 Db 904 AGCGGCGCGGTGACGCGAGCATGAGCCGCGAGACCAAGAGAGCGCACGCTCTTC 963  
 QY 258 ArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle 277  
 Db 964 AGGGTGAAGCGGTGACATACAGTACGCGCGCGCTCGGCGCACCATCACTCACTGTC 1023  
 QY 278 LeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnGluTyrTyrGlnSer 297  
 Db 1024 TTGGTCAAGAGACACCAACGACCAAGCCGCGCTTTCAGACAGTCCGAGTACCGCGAGCGC 1083  
 QY 298 LeuArgGlnAsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAsp 317  
 Db 1084 GTGCGGAGAACCTTGAGGGGTCTACAGAGTCTACACATCCGCGCGCACCGCGAGC 1143  
 QY 318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGluGlySerGlyLysSerProSer 337  
 Db 1144 TCGCCCATCAAGCGCAACTTGCGTTACCGGTGTTGGGGGGCGCGG----- 1191  
 QY 338 GluValPheGluValLeuAspProArgSerGlyValIleArgThrArgGlyProValAlaParg 357  
 Db 1192 GACGTCTTCAGGTCAACAGAGAGCTGTGGGTGTGAGACACGCGCGGTGTGAGCCG 1251  
 QY 358 GlnGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnIleArgAspProGly 377

Db 1252 GAGAGGCGCGCGAGTACCACTCTCTGTGAGGCGCAACGACGAGGCGCGCAATCCGCGC 1311  
 QY 378 ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAlaPro 397  
 Db 1312 CCGCTCATGCGCAGCGCGCACCGGTATCATCGAGGTGAGAGACGAGAACGACATCACTCC 1371  
 QY 398 GlnPheSerGluYleuArgTyrValValGlnValArgGluAspValThrProGlyAlaPro 417  
 Db 1372 CAGTTCAGCGCAGCAAACTACGTGTCCAGGTGCCGAGGAGCGTGGCGCTCAACACGCGCT 1431  
 QY 418 ValLeuArgValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHisTyrSer 437  
 Db 1432 GTGCTGAGATGACAGCGCACGAGCCGAGACCGAGACGAGAACCGCGCATTCATCAACAGC 1491  
 QY 438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp 457  
 Db 1492 ATCTCATGCGGAGAACGTGCGCGCGCAGTTCATCTGCACTCGCGAGCGGAGATCTGGAT 1551  
 QY 458 ValValSerProLeuAspTyrGluThrThrIleGluTyrThrLeuArgValArgAlaGln 477  
 Db 1552 GTGATCAACCCCTTGATTTTCAGAGATGTCAGAAATATCTGCTGAGCATTAAGCGCCAG 1611  
 QY 478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
 Db 1612 GATGGGCGCGCGCGCGCGCTCATCAATTTTCAGAGGAGGTGTGTGTGAGAGTGTCTGAT 1671  
 QY 498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSer 517  
 Db 1672 GTCAACGACACAGACCGTATTTGTGAGAGCGCCCTTCAGGCGCACGCTGCGAGAT 1731  
 QY 518 ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
 Db 1732 GTGCCCTCGGCGTACCCCGGTGTGACATTCAGCGCGGTGACGCGGACTCTGAGAGAAC 1791  
 QY 538 AlaArgLeuGluTyrArgLeuAla-----GlyValGlyHis--- 549  
 Db 1792 GCGCGGTGACATATCGCTGTGTGAGACGCGCTTCACCTTTCTGGGGGCGCGCGCT 1851  
 QY 550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db 1852 GGGCGTAAAGATCTGCGCGCGCGCACCGCTGACTTCCCTTCAGATTCACACAGCTCCGCT 1911  
 QY 561 TrpIleSerValAlaAlaGlnLeuAspArgGluValAspPheTyrSerPheGlyVal 580  
 Db 1912 TGGATTCACAGGTGTGCGGACTGACCGCGAGAGGTGAGCACTACAGCTTCGGGGTG 1971  
 QY 581 GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1972 GAGCGGTGAGACACGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGATCGAGTG 2031  
 QY 601 LeuAspValaAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 Db 2032 CTGAGACGTGAATGACAAAGACCGGTTGTCAGGAGCGCACCTAGAGCTTCGTGAT 2091  
 QY 621 GluAspAlaAlaValGlyThrSerValValThrValSerAlaValaAspArgAspAlaHis 640  
 Db 2092 GAGGATGGCGCGGTGGGAGCGGTGTGACTGACCTTCGAGCGCGCGCGAGCCGAGCGCAAC 2151  
 QY 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 2152 AGTGTGATTAATCACTACACTACAGCGCGCGCGCGCGCGCGCGCGCGCTTGTGACTGAGCGC 2211  
 QY 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluGln 680  
 Db 2212 CAGAGAGGGGCGCGCTCATCACTCGGCGCTACTCTGAGCTACAAAGAGAGAGAGAGCAG 2271  
 QY 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 2272 TACGTGTGGGGGTGACAGCATTCGAGCGGCAACGAGTGCACACTGCGCATGCTTAATC 2331  
 QY 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720

Db 2332 AACGTCACTGATGCGCAACACCCAGGCGCTGCTTTCAGAGCTCCATTACACGTGAGT 2391  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValIleu1SerAlaThrAspGluAsp 740  
Db 2392 GTCAAGTGAAGACAGCGCTGTGGGCACTTCATGTAACCTGATGCCAAGATGAGGAC 2451  
Qy 741 ThrGlyGluAsnAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACAGGAGAAATATGCGGCAATCACTGATGACAGACCCCGTGGCGGCACTTCCCAAT 2511  
Qy 761 AspAlaAspThrGlyValAlaValThrThrGlnAlaGluIleuAspThrGluAspGluIleu 780  
Db 2512 GACCCCGGACAGTGGGACCAATGACCAATGATGAGTGGACATGATGAGGACAGGTGCGC 2571  
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleuSerAspThrThr 800  
Db 2572 TACACCTGATCCATCAATGCGCCGAGCAACAGGCAATCCCGCAAAATCAAGACCAACAC 2631  
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
Db 2632 CTAGAGATCTCAATCTCGATGCAATGCAATGCAATGCCACCCAGTTCTGTGGGATTTCTAC 2691  
Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValIleuGlnIleSerAlaThr 840  
Db 2692 CAGGCTTCATCTTGAAGATGCTCAACCTCGACCAAGCATCTTCAAGGTCTGTCCACG 2751  
Qy 841 AsparGAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860  
Db 2752 GACCGGAGCTCAAGGTCCCAATGGGCGCTGTGCTGTACACCTTCAAGGTGGGGAGCGAGC 2811  
Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuArgIleuAsp 880  
Db 2812 GATGGGAGACTTCTTACATGAGCCAGCCAGTCCGGGTGATTCGCAACCGAGCGCGCTGGAC 2871  
Qy 881 ArgGluAsnValAlaGlnTyrValIleuArgAlaTyrAlaValAspIleuMetPro--- 899  
Db 2872 CCGGAGATGTGGCGCTGTCACCACTTGGGCTCTGGCTGTGATGGGCGAGTCCCACT 2931  
Qy 900 ProAlaArgThrProMetGluValThrValThrValIleuAspValAsnAspAsnProPro 919  
Db 2932 CCCCTTAGGCGCTCGGTGAAATCCAGGTGACCACTTGGACATTAATGACATGCCCCC 2991  
Qy 920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 ATGTTTGAAAGAGACCAATGAGCTGTGTGAGAGAGAACAAACCAAGTGGGCTGGTG 3051  
Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGCAAAATTCGTGCTAACACCTGATGAAAGCCCTAATGCCAATCATGATTCAG 3111  
Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyValIleuThr 979  
Db 3112 ATTTGTGAGAGGGAGCAATGCGGCACTTCTTCAGCTGGACTGTCAACGGGAGCTGGCT 3171  
Qy 980 AlaLeuValAspLeuAspThrGluAspArgProGluTyrValIleuValIleGlnAlaThr 999  
Db 3172 GCCATGTGTAGCTGACTTGTAGGTCGCGGAGATATGTGTGTGTGTGAGGCGACG 3231  
Qy 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuIleuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGCTGTGAGCGGAGCGGACGATGTCATCTTCTGTGACCAAGATGACAC 3291  
Qy 1020 ProProValIleuGlyAsnPheGluIleuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3292 CCGCGCTGTGCTGCGCACTTCAGATCTTCTCAACATATGTCACCAAGTCCACAC 3351  
Qy 1040 SerPheProGlyValAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGGCGGTATCGGCTGATCCGCGCCATGACCCCACTGTCCAGACAC 3411  
Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluIleuSerLeuValIleuAsnAlaSerThr 1079  
Db 3412 CTCACATTAACCTCTGTGAGGCGGCAACGAGCTGCGCTGTGTGTGTGACCCCGGACG 3471

Qy 1080 GlyGluLeuIleuSerArgAlaLeuAspAsnAsnArgProLeuGluValAlaIleMetSer 1099  
Db 3472 GCGAATCTGCACTCAGCGCGGAGCTGTGACCAACACCGGCGCTGGAGCGCTGATGAG 3531  
Qy 1100 ValIleuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 3532 GTGTGTGTGTGATGAGGATTCACAGCGCTCACGGCTTCTGTGACCTGTGTGTACACAC 3591  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgGluAspMetSerProGlu 1139  
Db 3592 ATCAAGAGACATGCTGATCCACAGCATCATCTGCTGCTGAGAACATGTCACGAG 3651  
Qy 1140 ArgPheLeuSerProLeuIleuGlyPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 3652 AAGTTCCTGTCCCGCTGTGGCGCTTCTGTGAGGGGGGTGGCGCGCTGTGTACAC 3711  
Qy 1160 ProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyValHis 1179  
Db 3712 ACCAAGAGACAGCTTGTGTCTTCAACGTCCAGAACGACACCGACGTC---AGCTCCAAC 3768  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyValGlyProProPhe 1199  
Db 3768 ATCTGAACGTACCTTCTCGGCGCTGCTGCTGGCGGCGTCCGCGC-----CAGTTTC 3822  
Qy 1200 LeuProSerGluAspLeuGlnGluArgLeuTyrIleuAsnArgSerLeuLeuThrAlaIle 1219  
Db 3823 TTCCTCGAGAGACCTCGACAGAGAGATCTACCTGAATCGGACCTGTGTACCAACATC 3882  
Qy 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Db 3883 TCCACGACGCGGTCTCCCTTCCACACCAACATCTGCTGGCGAGACCTCGGCGGCGCAAGAC 3942  
Qy 1240 TyrMetArgCysValSerValIleuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 3943 TACATGAAGTGTGTCTGCTGTGCAATTCAGACCTCGCGCTTCTCTCAAGCTCCACC 4002  
Qy 1260 SerValIleuPheArgProIleHisProValGlyIleuAsnArgCysArgCysProProGly 1279  
Db 4003 ACCGTGCTCTTCCGCGCATCCACCCATCAACCGGCGCTGCGCTGCGCGCGCGCGC 4062  
Qy 1280 PheThrArgIleAspThrCysGluThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 4063 TTCACCGCGACTACTGTGAGACGAGATGCACTCTGCTACTCCAGCCCGCGCGCGC 4122  
Qy 1300 HisGlyArgCysArgSerArgGlyValGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 4123 AACGCGCTCGCGGACGCGGAGCGGCGGCTACCTGCGAGTGTGAGAGACTTCACT 4182  
Qy 1320 GlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIleAsn 1339  
Db 4183 GGAAGACACTGTGAGGTGATGCGCGCTCAGGCGCTGTGCAACGGGGGTGTGCAAGAAC 4242  
Qy 1340 GlyIleThrCysValAsnLeuValGlyIlePheIleCysAspCysProSerGlyAsp 1359  
Db 4243 GGGGCACTCGGTGAACTGTCTCATCGGCGGCTTCCACTGCTGTGTCTCTCGGCGAG 4302  
Qy 1360 PheGluIleProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Db 4303 TATGAGAGGCTTACTGTGAGGTGACCAACAGAGCTTCCGCGGCACTTCTGTGTAC 4362  
Qy 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIleGlu 1399  
Db 4363 TTCGCGGCGCTGAGACAGCGCTTCACTTCACTTCCCTTCCCTCAAGTTGCACTGAGAA 4422  
Qy 1400 ArgAspGlyIleuLeuLeuThrThrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeu 1419  
Db 4423 AGAAGCGCTGTCTTCTTACAGCGCGCTTCAATGAGAGACAGACTTCACTGCGCGCTG 4482  
Qy 1420 GluValIleGlnGluValAlaGlnLeuThrPheSerAlaGlyIleuSerThrThrVal 1439  
Db 4483 GAGATCGTGAACAGAGAGTGTCAAGTCACTTCTGTGAGGAGAGACAAACAGACCGTG 4542



QY 1440 SerProPheValProGlyValSerAspGlyThrPheThrValGlnLeuValTyr 1455  
 DB 4543 GCACCAAGAGTCCAGTGTGTGAGTGAAGCGGCGGCGCATCTGTGACGGGCGAGTAC 4602  
 QY 1460 TyrAsnLeuProLeuLeuGlnThrGlyLeuProGlnGlyProSerGlnGlnVal 1479  
 DB 4603 TACACAAAGCCCAATTTGGCCACCTGGGCGCTGCCATGGGCGGCGGAAAGATG 4662  
 QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 DB 4663 GCGGTGTGACAGTGTGATGTGTGACAAACATGGCTGTGCGCTTGGAAAGGACATC 4722  
 QY 1500 GlyAsnTyrSerGlyAlaAlaGlnGlyThrGlnGlySerValGlySerLeuAspLeu 1519  
 DB 4723 GGGAACTACAGCTGCGCTGCGGCGACACAGCGGCTCCAAAGAGTCCCTGGAGTCCG 4782  
 QY 1520 ThrGlyProLeuLeuLeuGlyValProAspLeuProGlySerPheProValAspMet 1539  
 DB 4783 ACCGCGCTTACTCTTGGGGGGGTGTCCCACTGCGAGAACTTCCAGTGCACAAAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisLeuPheMetAla 1559  
 DB 4843 CGGCAAGTGTGGGCTGACAGCGAACTGTGACAGCGCAAAATGTGACATGGCC 4902  
 QY 1560 AspPheValAlaAsnAsnGlyThrValProGlyCysProAlaValValAsnValCysAsp 1579  
 DB 4903 GGAATATCGCCCAACATGTGACCCGGAAGCTGCGCTGCGAGAACTTCTCGCAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGln 1599  
 DB 4963 GGGAGCGGTGTGAGATGTGAGGCACTGTGTCAACAGTGTGAATGTATCTGTGTAG 5022  
 QY 1600 CysProLeuGlyPheGlyGlySerCysAlaGlnGlnMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTCCACTCGGATTCGGCGGGAAGAACTGTGACCAAGCATGCTCACCCCGCGCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaThrPheGlyLeuSerLeuProHisSerGlnProTyr 1639  
 DB 5083 AGCGGTGAGAGCTGTGCTGTGAGTGAACCTGAACATCATCTCTGTGCGCTGTGATC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaLeuThrArg 1659  
 DB 5143 CTGGGCTCATGTTCGCGACCCCGAAGAGAGACAGGCTTGTATGTGAGGCGACCAATGGT 5202  
 QY 1660 GlyArgSerThrLeuThrLeuGlnLeuArgGlyHisValMetLeuSerValGlnGly 1679  
 DB 5203 GGGCCACACAGCTTGGCTCCAGATCTGAAACAATCACTCCAGTTTGAAGGTGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTyr 1699  
 DB 5263 GCGCCCTCCGATGTGAGTCCGTGATGCTGTCCGGGTTCGGGTGTGACCGAGGGAGTGG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlnHisAlaIle 1716  
 DB 5323 CACCACTGTGTATCCAGCTGAAGAAATGTTAAGAGAGACAGTGAAGTGAAGCACTGGTC 5382  
 QY 1717 --LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValAsnLeuGlyProArgLeuHis 1735  
 DB 5383 ACCATGACCTTGGACTATGGAGTGAACAGAAACAAGGACAGATATCGGGGCGCATGCTCC 5442  
 QY 1736 GlyLeuHisValLeuSerAsnIleThrValGlyGlyTyrProGlyProAlaGlyValAla 1755  
 DB 5443 GGGCTAGCGGTGAAGAGCGTGTGTGAGGCGCTTGAAGACAAGCTTCCGTTGCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlnGlyValAlaAsn 1775  
 DB 5503 CGTGGATTCGAGAGCTCATGACAGAGAGTGAAGTGGGGGCGCCACCAACAGTCCGCC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
 DB 5563 ACCCTGAACATGAACAACGCACTCAAGTCAAGGTGAGAGCGGCTGTGTGTGTGAGGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815

DB 5623 CCTTACCTGAGGCGCTGTCCCGCAATAGCGGCTGCCAAGAGCGCTGGAGAGTAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 DB 5683 AGCTGCTGTGACAAAGGATTAAGTGAATTAATCTGTGAGAGCGCTGTCACTGAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgValProSerAlaProHisGlyTyrThrCys 1855  
 DB 5743 CCTCGAGAAACATGGGGGCGCTGCGTGCCTCCCGGCTCCCGACAGGCGTACCTGTGC 5802  
 QY 1856 GlyCysProProAsnTyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
 DB 5803 GAGTGTGGCCCAAGTCACTAGCGGCGGTGATCTGTGAACAAACTGACCTTCCGTCGCC 5862  
 QY 1876 ArgGlyTyrTrpGlyHisProThrCysGlyProCysAsnCysAspValSerLeuGlyPhe 1895  
 DB 5863 AGAGCTGTGTGGGGAACCCGTGTGTGACCTTGCACCTGTGCGGTGAGAAAGGCTTT 5922  
 QY 1896 AspProAspCysAsnLeuThrSerGlyGlyCysHisCysValGlnAsnHisTyrArgPro 1915  
 DB 5923 GATCCCGACTGTAAATGAACCAACGCGCAGTGCCTCAATGCAAGAAATTAATCAAGTTC 5982  
 QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 DB 5983 TAGCCCAAGACCTGTGTGCGCTGCGACATGTCTCCCATGTGCTCCACAGCGCGACT 6042  
 QY 1936 CysAspProGlnAspGlyGlnCysAspProCysValProGlyValIleGlyArgGlnCysAsp 1955  
 DB 6043 TGGACATGTGGCAACGGGCAAGTGTCTGCAAGCGCGGCGATCATGGCGGCAAGTGAAC 6102  
 QY 1956 ArgCysAspAsnProPheAlaGlyValThrThrAsnGlyCysGlnValAsnTyrAspSer 1975  
 DB 6103 CGCTCGCAACACCTTGTCCGAGGTGCACAGCTCGGCTGTGAAGTATCTACAGTGC 6162  
 QY 1976 CysProArgAlaIleGlnAlaGlyTyrTrpTrpProArgThrArgPheGlyLeuProAla 1995  
 DB 6163 TGTCCAAAGCATTTGAAGCGGCAATGTGTGGCCACAGACCAATTTGGGAGCGGCT 6222  
 QY 1996 AlaAlaProCysProLeuGlySerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 DB 6223 GCGGCGCATGTCCTAAGGAAATCGGTGAATGTGGTCCAGCATGCAAGCGGAGAAAG 6282  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlyLeuLeuGly 2035  
 DB 6283 GCTGTGCTGCGCCCAAGCTTTTAACGTATCAACACATCTCTGTGTGACCTCAAGGCG 6342  
 QY 2036 PheAlaGlyArgLeuGlnArgAsnGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 DB 6343 ATGAATGAAGATAGCGCGCATATGAACGAGTGAACGGGCGCAAGGCGCTGCAGCTG 6402  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValIys 2075  
 DB 6403 GTAGAGGCGCTGCGGAGTGTACACAGCAACGGGCAACGCTTGTGCAATGACGTGC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 DB 6463 AGGCGCTTCAAGCTGTGGGCGCAAGCTTCAAGCAAGAGCTGGCAGAGGGGCTTCAAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlnAsnLeuLeuArgValIleSerAlaLeu 2115  
 DB 6523 CTGGAGCGCAAGAGCGCGCACTTTCAGAGAGAGCGTATCATCTCGGAGCGCCCTTC 6582  
 QY 2116 LeuAspThrAlaAsnValArgHisTyrPheGlyLeuIleGlnThrGlnGlyTyrThrAla 2135  
 DB 6583 CTGGCGCCAGCAGCAGGCGCGGTGGAGAGCATGATCCAGCGGAGAGAGGCGGCGGCA 6642  
 QY 2136 TrpLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaAlaGlnAlaAsnMetArgHisThr 2155  
 DB 6643 CAGCTGTCTCGCGCTGAGGCGCTTACTTCAAGCAAGTGAACGGACCTGTGGCGGAGG 6702  
 QY 2156 TyrLeuSerProPheThrIleValThrProAsnIleValIleSerValValArgLeuAsp 2175







338 GUAUValPheGluValLeuAspProAspSerGlyValIleArgThrArgGlyProValAspArg 357  
1192 GAGCTTTCAGCTCAACAGAGAGCTGGCGTGTGTGACACAGCGCGGTGTGGACCCG 1251  
358 GUUUAUValGlySerGlyLeuThrValGluAlaSerAspGlnGlyArgAspProGly 377  
1252 GAGGAGCGCGCGAGTACAGCTCTCGTGTGAGGCCAACAGCCAGGGGGCAATCCGGGC 1311  
378 ProAspSerThrThrAlaValPheLeuSerValGluAspAspAspAspAspAspAsp 397  
1312 CCGCTCAGTGGCCACCGGCCAGCTGTACATCAGAGTGTGAGGAGAGAACAGCACTACCC 1371  
398 GlnPheSerGlnValArgGlyValValGlnValArgGluAspValThrProGlyAlaPro 417  
1372 CAGTTACAGGAGAGCACTACAGTGTGTCCAGTGTCCAGAGGACCTGTGGGCTCAACGCGCT 1431  
418 ValLeuArgValThrAlaSerAspArgAspGlySerAspAlaValIleArgThrSer 437  
1432 GTGCTCGAGTGCAGGCGCCACAGACCGGGACAGGGGCCAGAGAGCGGCCACTTCACTACCC 1491  
438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAsp 457  
1492 ATCTTCAGCGGGAGCGTGGCCGCGGCACTTCACTGTCACTCGCTGAGCGGGATCCGTGAT 1551  
458 ValValSerProLeuAspTyrGluThrTyrGlyGluTyrThrLeuArgValArgAlaGln 477  
1552 GTGATCAACCCCTTGATTTTCAGAGATGTCCAGAAATACCTCGCTGAGCACTTAAGGCCAG 1611  
478 AspGlyValArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
1612 GATGGGGGGCGCGCCCGCTCACTCAATCTTCAAGGGGTGTGTGTGTGTGTGTGTGTGTGT 1671  
498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
1672 GTCAACAGCAACAGCGCTATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1731  
518 ValProLeuGlyTyrLeuValLeuIleValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
1732 GTGCCCCGTGGGTACCCCGT 1791  
538 AlaArgLeuGluTyrArgLeuAla-----GlyValGlyHis--- 549  
1792 GCCCGGCTGCACCTATTCGCTGT 1851  
550 -----AspPheProPheThrIleAsnAsnGlyTyrGly 560  
1852 GGGCTTAAGAACTCTGCCGCCCAACCCCTGACTTCCCTTCCAGATCCAGCAAGCTCCGCT 1911  
561 TrpIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGlyVal 580  
1912 TGGATCACAGT 1971  
581 GlnAlaArgAspHisGlyTyrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
1972 GAGGGCGGTGAGCAACGCGCTCGGCCCAAGACTCTCTCAACAGCGTGTCTCACTACGCTG 2031  
601 LeuAspValAsnAspAsnAspProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
2032 CTGAGCGTGAATGACAGCAACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2091  
621 GlnAspAlaAlaValGlyTyrSerValValThrValSerAlaValAspArgAspAlaHis 640  
2092 GAGGATGCGCGCGT 2151  
641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAspArgPheSerIleThrSer 660  
2152 AGTGTGATTAACCTACAGCTCAACGCGGCAACCCCGAACCCTTGTGCACTACGCAACC 2211  
661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGluValGln 680  
2212 CAGAGAGGGGGCGGCGCTCATCAACCTGTGGCTACCTGTGACTCAACAGCAGAGCAGCG 2271

681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
2272 TACGTGTGGCGGTGACAGATCCCAACCGGTGCACACTGTGCATCTCTCAATC 2331  
701 AsnValIleThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
2332 AACGTCACTGATGCCAACCCCAAGCGCTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2391  
721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
2392 GTCACTGAGGACAGCGCTGT 2451  
741 ThrGlyGluAsnAlaArgGlyIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
2452 ACAGAGAGGAATGCCCGATCACTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2511  
761 AspAlaAspThrGlyAlaValThrThrGlnIleGluLeuAspTyrGluAspGlnValSer 780  
2512 GACCCCGACAGTGGACCACTGTACCACTGATGTAGCTGTGACTATGAGAACAGGTGTGCGC 2571  
781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
2572 TACACGTGACCATCATGTGGCCAGAGCAACCGCATCCCGAGAAATCAGACACCAACC 2631  
801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
2632 CTAGAGATCCCTCATCTCGATGCCAATGACATGACATGCCAGTTCCTGTGGGATTTCTAC 2691  
821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
2692 CAGGCTCATCTTGTGAGAGTCTCACCTTCAGACCAACATCTTCAGGTCTGTCCACG 2751  
841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860  
2752 GACCGGAGCTCAGGTCAGATGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2811  
861 AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgLeuAsp 880  
2812 GATGGGACTTCTCATGT 2871  
881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetPro--- 899  
2872 CCGAGAGATGTGGCGGTGTACACTTGTGGCTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2931  
900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspPro 919  
2932 CCCCTTACCGCGCTGT 2991  
920 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
2992 ATGTTTGAAGAGCAAGCACTGT 3051  
940 ValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGln 959  
3052 GTGGCAAGATTCGT 3111  
960 IleValGlnGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
3112 ATTGTGAGAGGAGCAGTCCGCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3171  
980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
3172 GCCATGT 3231  
1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAspAspAsn 1019  
3232 TCGCTCGCTGT 3291  
1020 ProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnAspSer 1039  
3292 CCGCTGT 3351  
1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059

Db	3352	AGTTTCCCAACGGGCGTGAATGGCGCTGATATCCCGGCCATATACCCGACGTGTGACACACG	34111
Qy	1060	LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuAsnAlaSerThr	10799
Db	3412	CTCAACTACACCTTCGTGACAGGGGCAACAGAGTGGCGCTGTGGCTGTGACCCCGCCACG	34719
Qy	1080	GlyGluLeuLeuLysLeuSerArgAlaLeuAspAsnAsnArgProLeuGluAlaGluMetSer	10939
Db	3472	GGCCAACTGCAGCTCAGCCCGGACCTGGACAACACCGGCGCTGAGAGGCGCTATGAGAG	35311
Qy	1100	ValLeuValSerAspGlyValAlaHisSerValThrAlaGluValAlaLeuArgValThrTle	11119
Db	3532	GTGTCTGTGTGTATGGATGATCACAAGCTCACGGCTTCCTGCACCTCGTGTGACCATTC	35919
Qy	1120	IleThrArgGluMetLeuThrHisSerTleThrLeuArgGluValAspMetSerProGlu	11339
Db	3592	ATCACGGACGACATGCTGACCAACAGCATCACTGTGCGCTGGAGAACATGTCCAGAGAG	36511
Qy	1140	ArgPheLeuSerProLeuLeuGlyLeuPheIleGluAlaValAlaAlaThrLeuAlaThr	11559
Db	3652	AAGTTCCTGTGCTCCGCGCTGTGGCCCTCTCGTGGAGGGGGGTGGCGCGGTGTGTCACAC	37111
Qy	1160	ProProAspHisValValValPheAsnValGluArgAspThrAspAlaProGlyGlyHis	11799
Db	3712	ACCAAGGACGACGCTTGTGTCTTCAACCTCCAGAACGACACCGACGTC---AGCTCCAC	37689
Qy	1180	IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyGlyProPhe	11959
Db	3769	ATCTTGAACTGACCTTCTCCGCGCTGTGCTGTGGCGCGCTGTCGGGCG-----CAGTTC	38229
Qy	1200	LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle	12199
Db	3823	TTCCCGTCGGAGGACCTGECAGAGAGAGATCTCACTGAATCCGAGACCTGCTGACCAACATC	38829
Qy	1220	SerIleGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluValProCysGluAsn	12339
Db	3883	TTCACGACGCGCGTGTGCTCCCTTCGACACACATCTGCTGTGGCGGACCTTGCGAGAAC	39429
Qy	1240	TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer	12559
Db	3943	TACATGAAGTCGATGTCCGTTCTGTGGATTTCAGACACTCCGCGCCCTTCTCAGCTCCAC	40029
Qy	1260	SerValLeuPheArgProIleHisProValGlyGlyLeuAspArgValTyrCysArgProProGly	12799
Db	4003	ACCGTGTCTTCCGCGCCATCCACCCCATCAACAGCGCTGCGCTGCGTCCCGCCCGCGC	40629
Qy	1280	PheThrGlyAspArgTyrCysGluThrGlyValAspLeuCysTyrSerArgProCysGlyPro	12939
Db	4063	TTTCAACCGCGCATCTAGTGGACAGCGAAATCAACTCTGTCTACTCCGACCCGTGCGGCGCC	41229
Qy	1300	HisGlyArgCysValArgSerArgGluGlyGlyTyrThrCysLeuCysValArgAspGlyTyrThr	13199
Db	4123	AACGCGCGCTCCGCGACCGGAGGGCGGCTTACACTCGAGTGGCTTTCAGGACATTTCACT	41829
Qy	1320	GlyGluHisCysGlyGluValSerValAspSerGlyArgCysThrProGlyValCysLeuAsn	13339
Db	4183	GGAAGACACTGAGGTGATGATGCCGCTGAGGCGCTGTGCCACAGGGGTGTGACAGAAC	42429
Qy	1340	GlyGlyThrCysValAsnLeuLeuValGlyGlyPheLysCysAspCysProSerGlyAsp	13559
Db	4243	GGGGGCACTCGTGAACCTGTGATCATCGGCGGCTTCSAATGCGTGTGCTCTCTGGAGAG	43029
Qy	1360	PheGlyLysProCysCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr	13799
Db	4303	TATGAGAGGCGCTTATGTGAGGTGACCAACAGAACTTCCGCGCCCAAGTCTTGTCTCAACC	43629
Qy	1380	PheArgGlyLeuArgGlyAlaArgPheHisPheThrLeuAlaLeuSerPheAlaThrLysGly	13939
Db	4363	TTTCGGGGGCGTGAAGCAAGCGCTTCACTTCACTCACTCCCTCAAGTTCAGGAA	44229
Qy	1400	ArgAspArgIleLeuLeuLeuTyrAsnGlyArgPheAsnGlyLysHisAspPheValAlaLeu	14199

Dd	4423	AGGAACGGCTTGCTTCTCTAACAGCGCGGCTCAATGAGACAGCACTTCAATCGCCTG	4482
Qy	1420	GIUValIleGInGInValGInLeuThr-PhseSerIaGlyIuSerThrThrVal	1439
Dd	4483	GAGATCGTGAACGAGCAGGTGCAGCTCACTTCTCTCAGCGCGAGAACACACACGCTG	4542
Qy	1440	serProPheValProGlyGlyValIserAspGlyGInTrpHisThrValGInLeuTyTyr	1459
Dd	4543	GCACCGAAGGTTCACATGGTGTGTAGTACCGGGCGGTGGCACTGTGTGCAGTGCAGTAC	4602
Qy	1460	TyrAnIyPProLeuLeuGlyGInThrGlyLeuProGInGlyProSerGInGInIyVal	1479
Dd	4603	TACAAACAAGCCCAATATATGGCCACCTGGGCGCTGCCATGGCGCTCGGGGAAAAATG	4662
Qy	1480	AlaValValThrValAspGlyCysAspThrGlyValAlaLeuMyArgPheGlySerValLeu	1499
Dd	4663	GCGGTGTGACAGTGTGATGTGTGACACACATGGCTGTGGCTTTGGAAAGACATC	4722
Qy	1500	GlyAnIySerCysValaIaGInGlyThrGInGlyGlySerIyIySerLeuAspLeu	1519
Dd	4723	GGAAATACAGCTGCGCTGCCACAGGCACTACACCGGCTCCAAAGATCCCTGATCTG	4782
Qy	1520	ThrGlyProLeuLeuGInGlyValProAspLeuProGlySerPheProValArgMet	1539
Dd	4783	ACCGCGCTTACTCTCCGGGGGGTGTCCCAACTGGCAAGAATCCCAAGTCACAC	4842
Qy	1540	ArgGInPheValGlyCysMetArgAsnLeuGInValAspSerArgHisIleAspMetAla	1559
Dd	4843	CGGCAATTCGTGGGCTGCATCCGGAACCTTGCACTGCACCGGCAAAATGTGTGACATGGCC	4902
Qy	1560	AspPheIleAlaAsnAnGlyThrValProGlyCysProAlaIyValyValAnValCysAsp	1579
Dd	4903	GGATTCATCGGCACAAATGGACCCGGGAAGGCTGCGCTGCGGAATCTTGCGAT	4962
Qy	1580	SerAnThrCysHisAsnGlyGlyThrCysValAsnGInTrpAspAlaPheSerCysGlu	1599
Dd	4963	GGAGAGCGGTGTCAAGATGGAGGCACTGTGTCAACAGCTGGAAATGTATCTGTGTAG	5022
Qy	1600	CysProLeuGlyPheGlyGlyIySerCysAlaGInGluMetAlaAsnProGInHisPhe	1619
Dd	5023	TGTCCACTCCCAATTCGGCGGGAGAACTGTGACGACACATGCTCCACCCCACTCTTC	5082
Qy	1620	LeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGInProTrpTyr	1639
Dd	5083	AGCGGTAGACGCTGCTGTCTGGAGTACCTTAACATCATCATCTGTGCTCTGGTAC	5142
Qy	1640	LeuSerLeuMetPheArgTrpArgGInAlaAspGlyValLeuLeuGInAlaIleThrArg	1659
Dd	5143	CTGGGGGCTCATGTTCCGGAACCCGGAGAGGACAGCGTTCTGATGTAGAGGCCACAGTGT	5202
Qy	1660	GlyArgSerThrIleThrLeuGInLeuArgGInGlyHisValMetLeuSerValGInGly	1679
Dd	5203	GGGGCCACACACTTTCGCTCCAAATCTTGAACAACTACCTCCAGTTGAGTGTCCAC	5262
Qy	1680	ThrGlyLeuGInAlaSerSerLeuArgLeuGInProGlyArgAlaAspGlyAspTrp	1699
Dd	5263	GGCCCTCCCATGTGTGAGTCCGTATGCTGTCCGGGTTCGGGTGACCGAGGGGAGTGG	5322
Qy	1700	HisHisValaGInLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle	1716
Dd	5323	CACCAACCTGCTGACGAGTGAAGAAATGTTTAAGAGAGACAGTGAAGACACTGTGC	5382
Qy	1717	---LeuSerPheArgTrpGlyGInGInArgAlaGInGlyLeuMetLeuGlyProGluGInHis	1735
Dd	5383	ACCATGACTTTGGACTTTGGATGTACACACAAACAGGACATATCGGGGGAGTCTTCC	5442
Qy	1736	GlyLeuHisIleLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla	1755
Dd	5443	GGGCTGACGGTAAGAGAGGTGTGTGTGGAGGCGCTCTGAAGACAAAGTCTCGGTGCG	5502
Qy	1756	ArgGlyPheAspGlyCysLeuGInGlyValArgValaSerAspTrpProGInGlyValAsn	1775
Dd	5503	CGTAGATTCGAGGTGTGATGACAGGATGTAGATGTAGGGGGAGGCCACCAAGTGTGCC	5562

Qy	1776	SeetleuapProserhiaglyleuSerlleasenValglunglmglyCysserleuProasp	1795
Dp	5563	ACCTGAACATGAACAGCACTCAAGGTCAAGGTGAAGAGACGGCTGTGATGTGAAGAC	5622
Qy	1796	ProCyaaPserAapProCyaaProAlaAnsSerTyCysserAanAapTPAapSerTyx	1815
Dp	5623	CCCTGTAACCTGAAGCCCTGTCCCCCAATAGCCGTCCACAGACGCTTGAGAGACTAC	5682
Qy	1816	SerCysserCyaaPProglyTytyrTglyaapAnsCythrAanValCyaaPleuAan	1835
Dp	5683	AGCTGCGCTGTGCAAGGGGTACTTGGAATTAACCTGTGGAATGCTGTCACTGAAC	5742
Qy	1836	ProCyagluhiagInservalCyethrAxyuPProserAlaProhiaglyTythrCy	1855
Dp	5743	CCCTGCGAAGAACATGGGGGCTGTGGTGGCTCCCGGCTCCCGCAAGGTACTAGTGGC	5802
Qy	1856	GlucyPProPoaenTytleuglyProTyCygluThzArglleaPglInProCyaaPro	1875
Dp	5803	GAGTGTGGGCGGACGACTACGGGCGGTACTGTGAACAACATCGACCTTCGGTGGCCC	5862
Qy	1876	ArgglyTyTPTrpglyhAProThCyaglyProCyaaAnsCyaaPValSerTyglyPhe	1895
Dp	5863	AGAGGCTGTGTGGGGAACCCGCTGTGTGAACCTGCACATGTGCGCTGACGAAGGCTTT	5922
Qy	1896	AapPProaspCyaaAnsTyThSerTylGluCyhiAsCylyuSgluaAnshtyArgPro	1915
Dp	5923	GATCCCACTGTATTAAGACCAACGCGCAAGTGCACATGCAAGGAATTTACTCAACATC	5982
Qy	1916	ProglySerProthrCyVleuLeuCyaaPcyAtyrProthrglySerleuSerArgVal	1935
Dp	5983	CTAACCCAGGACACTGTCTGTGCGGTGACACTGTCCCGCATGGCTGCCACAGCCGACT	6042
Qy	1936	CyaaPProgluaapgluInCyaaProCyaaPProgluVallleaglyArguInCyaaP	1955
Dp	6043	TGCACATGGCCACCGGAGCAGTGTGCTGCAAGCCCGGCTCATGTGGCGCATGTGAC	6102
Qy	1956	ArgCyaaPaaPProPhealagluValThrThzAanglyCyagluValAamTyaaPser	1975
Dp	6103	CGCTGCACAAACCGTTTGCAGAGTCAACACGCTCGGCTGTGAAGTATCAAAATGCG	6162
Qy	1976	CyaaPProarglallegluAaglylleTPTrProarghAarghAarghgluProAla	1995
Dp	6163	TGTCCCAAGCATTTGAAGCCGGGATCTGTGTGGCCACAGAACCAAGTTCCGGCAACCGGCT	6222
Qy	1996	AlaAlaProCyaaProLySglYserPheglyThzAlaValArghtsCyaaPgluhiArg	2015
Dp	6223	GCGGTGCATGCCCTTAAGGATCCGTTGTGAATGCCGTCCACACTGCACAGCGGGAAG	6282
Qy	2016	GlyTyTPLeuProPoaenLeuPheAnsCythrSerlleThrPheSergluLeuTygely	2035
Dp	6283	GCGTGGCTGCCCCCAGAGCTCTTTAACTGTACCAACATCTCTTCGTGAGACTCAAGGGCC	6342
Qy	2036	PhealagluArgleuGlnArgAangluserTyleuAapsergluYargsergluInleu	2055
Dp	6343	ATGAATAGAAGCTGAAGCCGCAATGAAGACGAGGTGAAGCGGCGCAGAGGCTTCGAAGT	6402
Qy	2056	AlaLeuLeuLeuAargAamaIaThGlnhiAThrAlaGlyTyrPheglySerAapValTy	2075
Dp	6403	GTAAGGCGGCTGCGAGCTTACACAGACAACGGGCAACGCTCTTTGGCAATGAACGTGCG	6462
Qy	2076	ValAlaTyrgluLeuAaIaThrArgleuLeuAlahtagInservalThrglnAargglyPhegly	2095
Dp	6463	ACGGCCTTACAGCTGTGGGCGCACTCTTGAACAAGAGGTGGCAAGCGGCTTCGAC	6522
Qy	2096	LeuSerAlaThrglnAapValhiAphetheThrgluAnsLeuLeuArgValgliserAlaLeu	2115
Dp	6523	CTGGCAGCGACAGAGAACGCCGACTTTCAAGAGGACGTATTCACCTGGGCAAGGCCCTC	6582
Qy	2116	LeuAapThzAlaAanLyAarghtsTyrgluLeuIleaglInthrgluTygluTyhAla	2135
Dp	6583	CTGGCCTCCAGCACAGAGCGGCGCTGTGAAGAGATTCAGAGCGAGCGGCGGCGGCA	6642

QY	2136	TrpleuLeuEnGI.nh1sYrYgUua1aTyZalAserAlaleuAlaGlnaMeeArgh1sThr	2135S
Db	6643	CAGCTGCTCCGCGCCCTCGAGGGCTACTTCAGCAACCGAGGACCGAACGGTCCGGCGAGCG	6702Z
QY	2156	TyrlseuSerProPheThr1leVal1ThrProaMileVal1IleSerVal1ValArgLeuAsp	2175S
Db	6703	TACCTGGGGCCCTTCGTCAATGTCACCCGCCAACATGATTCCTTGCTGTGACATCTTGAAC	6766Z
QY	2176	LysG1yAsnPhleuAlaG1yAlaLaleuLeuProArGtyrG1yAlaLaleuArG1yG1uGlnPro	2195S
Db	6763	AAGTTCACCTTACGGGAGCCAGGGATCCCGGATTCGACACATCCAGAAAGAGTTCCTCC	6822Z
QY	2196	ProaPleuEnGIuThrThVal1IleuProG1uSerVal1PheArgY-----G1u	2212Z
Db	6823	AGGAGCTCGAGTCTCCGCTCTCTCCACCGCACTTCTCAAGCACCTGMAAGAAAAA	6888Z
QY	2212	ThrProVal1ValArgProAlaG1y-----ProG1y	2222Z
Db	6883	GAAGGCCCCCTCGTAGGGCCGGCTGGCCCGAGAGACCAACCCCGACAGACCAACGCCCGGGG	6942Z
QY	2223	G1uAlaGlnG1uProG1uG1uLeuAlaArgAlaGlnArgh1sProG1uLeuSerGln	2242Z
Db	6943	CCTGGCACCGAGAGAGAGGCCCGCATCAGCAGGGGAGGACACCCCTGAAGACGTGGC	7002Z
QY	2243	G1yG1uAlaValAla1AserVal1Ile1sYrArGtyrThrlleuAlaG1yLeuLeuProH1sAsn	2262Z
Db	7003	CAGTTCCGCGCTCGCTCGTCACTATTCAGCACCCCTGGGGGAGCTCTCCCGAGGCC	7062Z
QY	2263	TyrAspProAspAluValArgSerLeuArgVal1ProLysArgProIle1leAsnThrProVal	2282Z
Db	7063	TACAGACCCCGACCGTCCGACGCTCCGGTGTCTCACCGGCCCACTAAATAAACCCGATG	7122Z
QY	2283	ValSer1leSerVal1H1sAspAspG1uG1uLeuLeuProArG1aLaleuAspLysProVal	2302Z
Db	7123	GTAGACACGCTGGTGTACAGGAGAGGGGCTCCGCTCCGAGACCCCTGGAGAGGCCGCTC	7182Z
QY	2303	ThrValGlnPheArgLeuLeuG1uThrG1uG1uArgThryrAspIleCyAsVal1PheTyr	2322Z
Db	7183	CTGGTGGAGTTCGCCCTCGTAGAGTGGAGGAGGAGAACCAAGCTGTCGTGGTTCGTGG	7242Z
QY	2323	AsnH1sSer1leLeuValSerG1yThrhG1yG1yThrSer1laArg1yCyAsG1uVal1Val	2342Z
Db	7243	AACCACTCCCTGCGCCGTGGTGGAGACGAGAGGTGTCTGTCCCGGGCTGGACATCTCTG	7302Z
QY	2343	PheArgAsnG1uSerH1sVal1SerCyG1uCyAsAsnH1sMetThrSerPheAlaValLeu	2362Z
Db	7303	TTCAGGAACCGGACACATGTCTGCTGCAGTGACGACACAGCCAGCTTTGGGTGTCTC	7362Z
QY	2363	MetLysVal1SerArGArgG1uAsnG1yG1u1leuLeuProLeuLysThrlleuThrTyVal	2382Z
Db	7363	ATGGAATATCTCCAGGCTGAGAAAGGGAGGTCTGTGCTCTGAAAGATGTCACTTAATGCC	7422Z
QY	2383	AlaLeuG1yVal1ThrlleuAlaAlaLeuLeuLeuThrPhePheLeuThrThrlleuLeuArg	2402Z
Db	7423	GCTGTGTCTTGTCACTGGCAGCCCTCTGTGTGCTCTTCGCTCCCTCGAGCTGTGTCCG	7482Z
QY	2403	1leLeuArgSerAsnGlnH1sG1yTleArgArgAsnLeuThrAlaAlaLeuG1yLeuAla	2422Z
Db	7483	ATGCTGGCTCCCAACCTGCACAGCATTCACAAGCACCTCCCGGGGCGTCTCTCTCTCT	7542Z
QY	2423	G1uLeuVal1PheLeuEnGIy1leAsnG1uAlaAspLeuProPheAlaCyAsThrVal1Ile	2442Z
Db	7543	CAGCTGTGTTCGTGATGGAGATCAACAGGAAACCGGTTCGTGTGACAGTGGT	7602Z
QY	2443	Ala1leLeuLeuH1sPheLeuTyrlleuCyAsThrPheSerTyrAlaLeuEnGIuAlaLeu	2462Z
Db	7603	GCCATCTCTCTCCACTACATCTACATAGACACCTTTCCTGAGACCTCGTGGAGAGCTTG	7662Z
QY	2463	H1sLeuTyrArgAlaLeuThrArg1uVal1ArgAspVal1AsnThrG1yProMetArgPheTyr	2482Z
Db	7663	CATGTCAACCGCATGCTGACCGAGGTGGCAACATTCACACGGGGCCATGCGGTTCTAC	7722Z
QY	2483	TyrlleuLeuG1yTyrG1yVal1ProAlaPhe1leThrG1yLeuAlaVal1G1yLeuAspPro	2502Z



```

Oy 57 uCyProSerSerAlaSerAsnLeuTyrThrSer----- 70
Db 169 AGCCCCCGGCGCGGAGCTGAGCGTGGCGGATGGCGGCTGAGACG 228
Oy 71 -----ArgCyAspAspAla-GlyThrGluLeuT 80
Db 229 CGCGCGTCTCGGCGCGGCGCGCGCTGCGCGCTGCAAGTCCGCTTGT 278
Oy 80 hrGlyHleValProHleAspGlyLeuArgValTyrCyProGluSerGluAlaH 100
Db 279 --GGCCCGAGTGGCC-----GAGCGGCTGAGCG-----CGCGTGGCGGCGCC 324
Oy 100 isIleProLeuProProAlaProGlu-GlyCyProTyrSerCyAspGluLeuGlyIle 119
Db 325 AC-----GCACCTTCCCGGCTGGAGCGCGCGCTGCGCGCTGCGGAAAC 369
Oy 120 GlyGlyHleLeuSerProGlnGlyLeuThrLeuPro----- 132
Db 370 GGTGGCCCGGCTTGC-----GGGCGCTTGTCTTCCCGTCCCGGCGGCTGCGCGCC 423
Oy 133 GlnGlnHleProCyAspLeuLeuAla----- 140
Db 424 GGCAGACATTCCGCGCTCGAGCTCGACACACCTTACCGGCTGCGCGCGCGCGC 483
Oy 141 ProArgLeuArgCyAspGlnSer----- 147
Db 484 CCGAGCGCCCGCTGTCCCGCGCTCCCATCTGCTGCGCGCGCGCGCTCGGTCGCGCTG 543
Oy 148 -----CyLeuLeuAlaGlnAlaProGly-----LeuArgAlaGly----- 159
Db 544 CGTCTGCTGTGCGCGCTGCGCGCGCGCGCTGCGCGCGCTGCGAGTGGAGCTGGAG 603
Oy 160 ----- 161
Db 604 GCGCGCACCGCGGAGCGCGCTCCGCGTGCATCCCATGCGCGCGCGCGCGCGCGAC 663
Oy 162 SerProGlnGlnSerLeuGlyGlyArgGlyValArgAsnValAsnThrAlaPro----- 179
Db 664 TTGCCCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGG 723
Oy 180 -----GlnPheGlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGly 197
Db 724 AGCGTGAAGTTCCAGTCCCACTACACAGCGTGGCGCTTGTGAAGACACACCGCGGCG 783
Oy 198 ThrProValAlaSerLeuArgAlaIleAspProAspGlnGlyGluAlaGlyArgLeuGlu 217
Db 784 ACCCTCATCTCCAGCTGACGCGACGACCTACACATCGAGGAGGAGAGAGGCGCTGAGC 843
Oy 218 TyrThrMetAspAlaLeuPheAspSerArgSerAsnGlnPhePheSerLeuAspProVal 237
Db 844 TATTACATGAGAGGCGCTGTTCACACGAGCGCTCCGCGGCTACTTCCGAATCGACTCTGC 903
Oy 238 ThrGlyAlaValThrThrAlaGlnGluLeuAspArgGluThrIleSerThrHleValPhe 257
Db 904 AGCGCGCGCTGAGACCGACAGCGCTGACCGCGAGACCAAGAGAGACGACCTCTC 963
Oy 258 ArgValThrAlaGlnAspHleGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle 277
Db 964 AGGGGGAAGCGCTGACTACAGTACGCGCGCGCGCTGCGCACACACATCACTCTCTC 1023
Oy 278 LeuValThrAspThrAsnAspHleAspProValPheGlnGlnGlnGlnTyrIleGluSer 297
Db 1024 TTGGTCAAGAAGACCAACGACGACCGCGCTTCCGAGCAGTCCGAGATACCGGACGCGC 1083
Oy 298 LeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAsp 317
Db 1084 GTGCGGAGAGACCTGAGGTGGCTACAGAGTGTGACCTACCGCGCGACGACCGCGAC 1143
Oy 318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyIleSerProSer 337
Db 1144 TCGCCCATCAACGCGCACTTGGCTTACCGCGGTGTGGGCGGCGCGTGG----- 1191
Oy 338 GlnValPheGlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArg 357

```

```

Db 1192 GACGCTTCCAGCTCAACGAGAGCTCTGCGGTGGAGACACGCGCGCGTGTGACCGG 1251
Oy 358 GlnGluValGlnSerTyrGlnLeuThrValGlnAlaSerAspGlnGlyArgAspProGly 377
Db 1252 GAGAGGCGCGCGAGTACCACTCTGTGTGAGGCGCAACGACCAAGGCGCGCTCGGCGC 1311
Oy 378 ProArgSerThrThrAlaIleValPheLeuSerValGlnAspAspAsnAspAlaPro 397
Db 1312 CCGCTCATGTCCACCGGCGACCGGTGTACATCGAGGTGAGAGAGAGAGACACATCACTACC 1371
Oy 398 GlnPheSerGlnValArgTyrValValGlnValArgGluAspValThrProGlyAlaPro 417
Db 1372 CAGTTCACGAGAGACAACTACGTGTCCAGGTGCCGAGGACGCTGGGCTCAACCGGCT 1431
Oy 418 ValLeuArgValThrAlaSerAspArgAspArgGlySerAsnAlaValAlaHleGlySer 437
Db 1432 GTGCTGCAGATGACAGGCCACGACCGGAGCACAGGCGCAAGACGCGCATTCATCAACGC 1491
Oy 438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValaLeuAsp 457
Db 1492 ATCTCATCGGGAACGTGGCGCGCACTTCTACCTGCACTGCTGAGCGGATCTGGAT 1551
Oy 458 ValValSerProLeuAspTyrGluThrThrIleGlnGlyTyrThrLeuArgValArgAlaGln 477
Db 1552 GTGATCAACCCCTTGATTTGAGAGATGTCCAGAAATCTCGTGAAGCATTAAGCCACG 1611
Oy 478 AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497
Db 1612 GATGGGCGCGCGCGCGCTCATCAATTTCCAGGGGTGTGTGTGTGTGTGTGTGTGTGTGT 1671
Oy 498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517
Db 1672 GTCAACAGACACAGACCTATCTTTGTGAGCAGCCCTTCCAGGCCACCGCTGAGAAAT 1731
Oy 518 ValProLeuGlyTyrLeuValLeuHleValGlnAlaIleAspAlaAspAlaGlyAspAsn 537
Db 1732 GTGCCCTCGGCGTACCCCGTGTGTCACTTCAGGCGGTGTGTGTGTGTGTGTGTGTGTGT 1791
Oy 538 AlaArgLeuGlnTyrArgLeuAla-----GlyValGlyHle--- 549
Db 1792 GCGCGCTGACATATCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1851
Oy 550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560
Db 1852 GGGCTTAAGATCTGCGCCCGCACCCCTGACTTCCCTTCCAGATCTCAACACGCTCGGT 1911
Oy 561 TrpIleSerValAlaAlaGlnLeuAspArgGlnGluValAspPheTyrSerPheGlyVal 580
Db 1912 TGGATCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1971
Oy 581 GlnAlaArgAspHleGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600
Db 1972 GAGCGGTGAGACAGGCTGTGCGCCCATAGCTCTTCCACACCGTGTCTATCAGGTG 2021
Oy 601 LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlnTyrThrValArgLeuAsn 620
Db 2032 CTGACCGTGAATGACACAGACCGGTGTTCACGAGCGCACCTACGAGCTTGTGTGAT 2091
Oy 621 GlnAspAlaAlaValGlyThrSerValValThrValSerAlaValaAspArgAspAlaHis 640
Db 2092 GAGAGTACGCGCGGTGGAGACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2151
Oy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660
Db 2152 AGTGTGATTAATCAACGCTCAACGCGGACACCGCGAAGCCCTTGTGACTAGAGCGC 2211
Oy 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnArgGln 680
Db 2212 CAGAGAGGCGCGCGCTCATCACCTGTGCGCTTGTGACTACAGAGAGAGACAGAG 2271
Oy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700

```



Db 2272 TACGTGCGGCGTGAACGATCCGACGCGTGCACATCGCGCATCTTCAATC 2331  
QY AsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisArgThrValAsn 720  
Db 2332 AACGTCACTATGCAACACCGACCGCTGTCTTTCAGAGCTCCATTCACACAGTAGT 2391  
QY ValAsnGluAspArgProAlaGlyThrThrValValLeuLeuLeuSerAlaThrAspGluAsp 740  
Db 2392 GTACAGTAGGACAGCGGCTGTGGGACCTCATTTGCTTACGCTGCGACATGAGAGAC 2451  
QY ThrGlyGluAsnAlaArgGlyLeuThrPheMetGluAspSerLeuProGlnPheArgGlyLe 760  
Db 2452 ACAGAGAGATGCGCGCATCACCTACGTATTCAGAGATCCCGCGCGATTCGGCATTC 2511  
QY AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspArgGluAspGlnValSer 780  
Db 2512 GACCCGACAGTGGCGACCATGTACACATGATGAGCTGAGCTATGAGAACCGAGTGGCC 2571  
QY TyrThrLeuAlaLeuThrAlaArgAspAsnGlyLeuProGlnAspSerThrThrTyr 800  
Db 2572 TACACGCTGACCATCATGCGCCAGGACACGCGATCCCGCAAAATCAGACACACACACC 2631  
QY LeuGluLeuLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspArgSerTyr 820  
Db 2632 CTAGAGATCTCTCATCTCGATGCGCAATGACATGCACTCCAGTTCCTGTGGATTTCTAC 2691  
QY GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnLeuSerAlaThr 840  
Db 2692 CAGGGTTCATCTTGAAGAGTCTCCACCTCGACGACGATCTCCAGGCTCTGCGCACG 2751  
QY AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyLeuAspArgGly 860  
Db 2752 GACCGGAGCTCAGAGTCCCATATGGGCGCTGCTGTACACTTCCAGGGTGGGAGACGAGC 2811  
QY AspGlyAspPheLeuValGluSerThrSerGlyLeuValArgThrLeuAspArgLeuAsp 880  
Db 2812 GATGGGAGCTTCTTACATCGAGCCGACGCTCGGTGTGATTCGACCCGCGCGCTGGAGC 2871  
QY ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspArgGlyMetPro--- 899  
Db 2872 CGGAGAGATTTGGCCGCTTACACCTTTGGGCTCTGCTGTGATCGGGCAATCCCACT 2931  
QY ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
Db 2932 CCCCTTAGCGCTCGGTAGAAATCCAGGTGACCATTTGACATTAATGACATATGCCCC 2991  
QY ValPheGlnGluAspGluPheAspValPheValGluGluAsnSerProLeuGlyLeuAla 939  
Db 2992 ATGTTTGAAGAGACGAACTGAGCTGTGTGAGAGAGAAACAACCGAGGGGTGCGTG 3051  
QY ValAlaArgValThrAlaThrAspProAspGlyGlyThrAsnAlaGlnLeuMetTyrGln 959  
Db 3052 GTGGCAAAAGTTCGTCTTACGACCTGATGAGGCCCTTAATGCCGATCATGTATCAG 3111  
QY IleValGlnGlyAsnLeuProGluValPheGlnLeuAspPhePheSerGlyLeuLeuThr 979  
Db 3112 ATTGTGAAGAGGACATGCGGCATTTCTTCACTGACCTGCTCAACGCGGAGCTGCGT 3171  
QY AlaLeuValAspLeuAspArgTyrGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
Db 3172 GCCATGTGAGCTGAGCTTGAAGCTCCGGCGGAGATGTGTGTGGTGGCGAGCCACG 3231  
QY SerAlaProLeuValSerAlaAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCGCTGGTGGAGCCGAGCCAGCGTGCATCTTCTCGTGAACCAAGATGACAC 3291  
QY ProProValLeuGluAsnPheGlnGlyLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
Db 3292 CCGCGTGTGCTGCGCGCATTCGAGATCTCTTCAACAATATGTCACCAAGATGCACAC 3351  
QY SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspPheSerAspSer 1059  
Db 3352 AGTTTCCCAACCGCGGTGATGGCTGATCCGCGCCATGACCCCGAGGTGACAGACGC 3411

QY LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1060  
Db 3412 CTCACCTACACCTTGTGTGAGGAGCAACAGCTGCGCTTGTCTGTGAGACCCCGCCACG 3471  
QY GlyGluLeuValLeuSerArgAlaLeuAspAsnAspArgProLeuGluAlaLeuMetSer 1080  
Db 3472 GCGAACTGCACTGACCGCGGACCTGAGCAACAACCGCGCTGAGGCGCTCATGAGAG 3531  
QY ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 3532 GTGTCTGTCTGATGAGCATTCACAGCGCTCAGCGCTTCTGACCTGTGTGTTCACATC 3591  
QY IleThrAspGluMetLeuThrHisSerLeuThrLeuArgGluGluAspMetSerProGlu 1139  
Db 3592 ATCAGGAGACATGCTGACCAACAGATCATCTGCTGCGCTGGAACATGTCCAGAGAG 3651  
QY ArgPheLeuSerProLeuLeuGlyLeuPheLeuGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 3652 AAGTTCTGTCCCGCTGCGCTCTTCTGTGAGAGGGGTGGCGCGCTGTCTTCACAC 3711  
QY ProProAspHisValValValPheAsnValGlnAlaAspThrAspAlaProGlyGlyHis 1179  
Db 3712 ACCAAGAGACAGCTCTTGTCTTCAACGTCGCAAGACGACGACGAGCTC---AGTCCAC 3768  
QY IleLeuAsnValSerLeuSerValGlyGlnProProGlyGlyGlyProPhePhe 1199  
Db 3768 ATCTGAACTGACCTTCTGCGCGCTGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCG 3822  
QY LeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAlaIle 1219  
Db 3823 TTCCTGCGAGAGACCTGACGAGAGACAGATCTTCAATTCGACCTGCTGACACATC 3882  
QY SerAlaGlnArgValLeuProPheAspAsnIleCysLeuArgGluProCysGluAsn 1239  
Db 3883 TCCAGCAGCGGTGTGCTGCTTTCACACCAATCTGCTGCGAGCGCTGCGAGCGCGAGAC 3942  
QY TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
Db 3943 TACATGAGTGTGTTCGCTTGTGATTCGACACTCGCGCGCTCTCTCAGCTCACAC 4002  
QY SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
Db 4003 ACCGTGCTTCTCCCGCCCATTCACCCCATCAACGCGCTGCGCGCTGCGCGCGCGCG 4062  
QY PheThrGlyAspTyrCysGlnThrGluValAspLeuCysTyrSerArgProCysGlyPro 1299  
Db 4063 TTCACCGGAGCTACTGCGAGAGAGATCGACTGTCTACTCCGACCCGTCGCGCGC 4122  
QY HisGlyArgCysArgSerArgGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
Db 4123 AACGGCGGTGCGGAGCGGAGCGGAGCGGCTTACACTGAGATGCTTGAAGACCTTCACT 4182  
QY GlyGluHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysValAsn 1339  
Db 4183 GAGAGCACTGTGAGGTGATGCCGCTCAGCGCCCTGTGCCAAGGGGTGTGCAAGAAC 4242  
QY GlyGlyThrCysValAsnLeuLeuValGlyGlyPheCysArgCysProSerGlyAsp 1359  
Db 4243 GGGGGACCTGCGTAACTGCTCATTCGCGGCTTCCACTGCGTGTCTCTTGGCGAG 4302  
QY PheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
Db 4303 TATGAGAGGCTCTTACGTGAGGTGACCAACAGAGACTTCCCGCCCAAGCTTCTGTACAC 4362  
QY PheArgGlyLeuArgAlaArgPheHisPheThrLeuAlaLeuSerPheAlaThrArgGlu 1399  
Db 4363 TTCGGGGCTTGAACAGGCTTTCACATTCACATCTCTTCACTGTTCCTCACTAGGAA 4422  
QY ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeu 1419  
Db 4423 AGGAACGCTTCTCTTACACAGCGCGCTTCAATGAGACAGACTTCACTCCCGCTG 4482

QY 1420 GlnValIleGlnIleGlnValGlnLeuThrPheSerAlaGlyGlySerThrThrThrVal 1439  
 DB 4483 GAGATCTGACGACGAGCTGACGCTCACTTCTCTGACGAGCAACAACGACGCTG 4542  
 QY 1440 SerProPheValProGlyGlyValSerAspGlyGlnIlePheValGlnLeuValTyr 1459  
 DB 4543 GCACCGAAGGTTCCCACTGAGTGAAGACGAGGAGGACCTCTGACGAGGACTAC 4602  
 QY 1460 TyrAsnLeuProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnVal 1479  
 DB 4603 TACACAGACGCCAATATTGGCACCTGGGCTGCCCATGGCCGCTCGGGAAGATG 4662  
 QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
 DB 4663 GCGGTGTGACAGTGAATGATGACACAACTGCTGCTGGCAAGAGACATC 4722  
 QY 1500 GlyAsnTyrSerCysAlaAlaGlnGlyThrGlnGlyGlySerValSerLeuAspLeu 1519  
 DB 4723 GGGAACTACAGCTGCGCTGCCAGGACCTGACCGGCTCCAAAGATCCCTGGATCTG 4782  
 QY 1520 ThrGlyProLeuLeuGlyGlyValProAspLeuProGlySerPheProValArgMet 1539  
 DB 4783 ACCGGCCCTTACTCTCGGGGGGTGCTCCCACTGACAGAGCTCCCACTGACAC 4842  
 QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
 DB 4843 CCGCAGTCTGTGGCTGTCACGCGAACCTGTCACTGCAACGCAAAATGTGACATGGCC 4902  
 QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaValSerValCysAsp 1579  
 DB 4903 GGAATCATCCCAACATAGGACCCGGAAGCTGCGCTCTCGAAGAACTTCTGGAT 4962  
 QY 1580 SerAsnThrCysHisAsnGlyGlyThrCysValAsnGlnTrpAspAlaPheSerCysGlu 1599  
 DB 4963 GGGAGCGGTGTGACATGAGGACCTGTGTCAACAGTGAAATGATCTGTGTGAG 5022  
 QY 1600 CysProLeuGlyPheGlyGlyValSerCysAlaGlnIleMetAlaAsnProGlnHisPhe 1619  
 DB 5023 TGTTCATCTCGATTCGCGGGAAGAACTGTGACAGAACCTGCTCCACCCCGACTCTTC 5082  
 QY 1620 LeuGlySerSerLeuValAlaTrpHisGlyLeuSerLeuProIleSerGlnProTrpTyr 1639  
 DB 5083 AGCGGTGAGAGCTCGTGTCTGAGTACCTGAACATCACTCTGTGCTGCTGATC 5142  
 QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
 DB 5143 CTGGGGCTCATGTTCCGAGCCCGGAAGAGAGACAGCTTCTGATGAGGCCACCAAGCT 5202  
 QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlyHisIleValMetLeuSerValGlyGly 1679  
 DB 5203 GGGCCCAACGCTTCCCTCGCATCTGACAACTGACAACTCACTTGAAGTGGCCAC 5262  
 QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTrp 1699  
 DB 5263 GGGCCCTCCGATGTGAGTCCGATGCTGTCCGGGTGCGGGTACCGAGGGAGATG 5322  
 QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
 DB 5323 CACCACTGCTGATCGAGTGAAGAAATGTTAAGAGAGACAGTGAAGTGAAGCACTGCTC 5382  
 QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnGlyValLeuLeuGlyProArgLeuHis 1735  
 DB 5383 ACCATGACCTTGAATATGAGATGAGACAGACAGACAGACAGATTCGGGGGCATGCTTCCC 5442  
 QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
 DB 5443 GGGGTACGTAAGAGACGCTGTGTGAGAGGCGCTCTGAAGACAAAGCTCCGAGCGC 5502  
 QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspTrpProGlyGlyValAsn 1775  
 DB 5503 CTGGGATTCGAGAGCTGACAGAGGAGTGAAGTGGGGGAGACGCCCAACAGCTGCGC 5562  
 QY 1776 SerLeuAspProSerHisGlyGlySerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795

DB 5563 ACCGTAAATGAACAACCACTCAAGGTGAAGAGACGGCTGTGATGTGACGAC 5622  
 QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyr 1815  
 DB 5623 CCTGTACTCGAGCCCTGTCTCCCAATAGCCGCTGCAACAACCTGGAGGACTAC 5682  
 QY 1816 SerCysSerCysAspProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsn 1835  
 DB 5683 AGCTGCTGTGTGACAAAGGTAACCTTGAAATTAATCTGTGTGATGCTGTCACTGAAC 5742  
 QY 1836 ProCysGlnHisGlnSerValCysThrArgLysProSerAlaProHisGlyTyrThrCys 1855  
 DB 5743 CCTGCGAAGACATGGGGGCTGCGGCGCTCCCGGCTCCCGCAGGGCTACGTGTGC 5802  
 QY 1856 GlnCysProProAsnTyrLeuGlyProTyrCysGlnThrArgIleAspGlnProCysPro 1875  
 DB 5803 GAGTGTGGCCCACTACATACGAGCGGTGTGTGAAACAACTGACCTTCCGTGCCCC 5862  
 QY 1876 ArgGlyTrpTrpGlyHisProThrCysGlyProCysAsnCysAspValSerLysGlyPhe 1895  
 DB 5863 AAGAGCTGTGGGGAAACCCCTGTGTGACCTTGCACATGTGCTCAGCAAGGCTTT 5922  
 QY 1896 AspProAspCysAsnLysThrSerGlyGlnCysHisCysAlaGlnValHisIleTyrArgPro 1915  
 DB 5923 GATCCCGACTGTATTAAGACCAACGGCCAGTGCACATGACAGAGAAATTACTACAGCTC 5982  
 QY 1916 ProGlySerProThrCysLeuLeuCysAspCysTyrProThrGlySerLeuSerArgVal 1935  
 DB 5983 CTAGCCCAAGACCTGTGTGCTGTGACCTGTCCCTCCCAATGAGCTCCCAAGCCGACT 6042  
 QY 1936 CysAspProGlyAspGlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAsp 1955  
 DB 6043 TGGACATGAGCCACCGGAGAGTGTGCTGTCAACCCGGGTATCGCGGCCAGTCCAAC 6102  
 QY 1956 ArgCysAspAsnProPheAlaGlnValThrThrAsnGlyCysGlnValAsnTyrAspSer 1975  
 DB 6103 CCGTGTGACAAACCGTGTGCTGACAGTCAACAGCTCGCGCTGTGAAGTATCAATATGCG 6162  
 QY 1976 CysProArgAlaIleGlnAlaGlyIleTrpTrpProArgThrArgPheGlyLeuProAla 1995  
 DB 6163 TGTCCAAAGCAATTGAGCGCGCACTGTGTGCGCAACAACAAGTTCGGAGCCGGCT 6222  
 QY 1996 AlaAlaProCysProLysGlySerPheGlyThrAlaValArgHisCysAspGlnHisArg 2015  
 DB 6223 GGGGTGCATGTGCTTAAGGATCTCGTGAATGCGGTCCACACTGACGCGGAGAG 6282  
 QY 2016 GlyTrpLeuProProAsnLeuPheAsnCysThrSerIleThrPheSerGlnLeuLysGly 2035  
 DB 6283 GCGTGGCTGCCCCAGAGCTTTTAACGTGTACCAACATCTCTTGTGTGACCTCAGGGCC 6342  
 QY 2036 PheAlaGlnArgLeuGlnArgAsnGlySerGlyLeuAspSerGlyArgSerGlnGlnLeu 2055  
 DB 6343 ATGAATGAAGACTAGCCGCAATAGACAGAGTGAAGCGCCAGCGCCCTGTGACGTG 6402  
 QY 2056 AlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValLys 2075  
 DB 6403 GTGAGGGCGCTGCGAGTCTACACAGACAGAGGCAAGCTTTGGCAATGACGTGCC 6462  
 QY 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlySerThrGlnArgGlyPheGly 2095  
 DB 6463 ACGGCTTACAGCTCTGCGCCACGCTCTTCAAGCACAGAGCTGACAGAGGGCTTGCAC 6522  
 QY 2096 LeuSerAlaThrGlnAspValHisPheThrGlyAsnLeuLeuValGlySerAlaLeu 2115  
 DB 6523 CTGGAGCCACCAAGAGCGCCACTTTACAGAGAGTATCACTCGGAGCGCGCTC 6582  
 QY 2116 LeuAspThrAlaAsnLysArgHisIleTrpGlyLeuLeuGlnGlnThrGlnGlyValAla 2135  
 DB 6583 CTGGGCCCAAGCAACGAGCGCGCTGTGAGAGATCCAGCGGAGAGAGGGCGGCAACGGA 6642  
 QY 2136 TrpLeuLeuGlnHisTyrGlnAlaTyrAlaSerAlaLeuAlaGlnAspMetArgHisIleThr 2155

Db 6643 CAGCTGCTCCGGGCTCGAGGGCTACTTCAGACGTCGACGCAACCTGCGCGAGACG 6702  
 Qy 2156 TyLeuSerProPheThrIleValIThrProAsnIleValIleSerValIValArgLeuAap 2175  
 Db 6703 TACCTCGGCGCTTTCGATCGTCACCGCCACATGATTTCTGCTGCACATCTTGGAC 6762  
 Qy 2176 TyGlyAsnPheAlaGlyAlaIleuProArgTyGlyIleuAlaLeuArgIleuGlnPro 2195  
 Db 6763 AACCTCAACTTTCAGGAGGACGAGGTCCTCCGATTCGACACATCATGAGAGTTCCTCC 6822  
 Qy 2196 ProAspLeuGlnThrValIleIleuProGluSerValPheArg-----Glu 2211  
 Db 6823 AGGAGCTGAGTCTCCCTGCTCTTCCACCGCACTTCTTCAGACCACCTGAAGAAAAA 6882  
 Qy 2212 ThrProValIValArgProAlaGly-----ProGly 2222  
 Db 6883 GAGGCGCCCTGCTGAGGCGCGCTGCGGAGACACCCCGGACACACCGCCCGGG 6942  
 Qy 2223 GluAlaGlnIleuProGluIleuAlaArgArgIleuArgIleuSerGln 2242  
 Db 6943 CCTGACACCGAGAGGAGGCGCCGATCAGACGAGGAGGACACCTGATGACGCTGGC 7002  
 Qy 2243 GlyIleuAlaValIleSerValIleIleTyArgThrIleuAlaGlyIleuProHisAan 2262  
 Db 7003 CAGTTCGCGCTGCTGCTGATCATATTCACGACCCCTGGGCGAGCTCTGCGCGAGCC 7062  
 Qy 2263 TyAspProAspIleuArgSerIleuArgValIProIleuArgProIleuIleuAnthrProVal 2282  
 Db 7063 TACGACCCGACCGGACGCTCCGCTGCTCACCAGCCCATCATTAATACCCGAGT 7122  
 Qy 2283 ValSerIleSerValHisAspAspGluIleuIleuProArgAlaLeuAspIleuProVal 2302  
 Db 7123 GTGAGACGCTGCTGATACAGCGAGGGGCTCCGCTCCGAGACCCCTGAGAGGCGCTC 7182  
 Qy 2303 ThrValGlnPheArgIleuIleuGlnIleuArgIleuArgIleuProIleuIleuPheTrp 2322  
 Db 7183 CTGCTGAGATTCCTCCCTGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7242  
 Qy 2323 AsnHisSerIleuValSerGlyIleuGlyIleuGlyIleuGlyIleuGlyIleuGlyIleu 2342  
 Db 7243 AACCACTCCCTGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7302  
 Qy 2343 PheArgAsnIleuSerHisValSerGlyGlnCysAsnHisMetThrSerPheAlaValIleu 2362  
 Db 7303 TCCAGAGACCGGACACATGCTGCTGACATGACGACACACACGCTTGGGCTGCTC 7362  
 Qy 2363 MetAspValSerArgArgIleuAsnGlyIleuIleuProIleuIleuIleuIleuIleu 2382  
 Db 7363 ATGGATATCTCCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7422  
 Qy 2383 AlaLeuGlyValIleuValIleuAlaIleuIleuIleuIleuIleuIleuIleuIleu 2402  
 Db 7423 GCTGTGCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7482  
 Qy 2403 IleuIleuArgSerAsnGlnHisGlyIleuArgAsnIleuIleuIleuIleuIleuIleu 2422  
 Db 7483 ATGCTCGCTCCCAACCTGACACATTCACACGACACCTCCGCTGCGCTCTCTCTCT 7542  
 Qy 2423 GlnIleuValPheIleuGlnIleuIleuGlnIleuAlaAspIleuProPheAlaCysThrValIle 2442  
 Db 7543 CAGCTGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7602  
 Qy 2443 AlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2462  
 Db 7603 GGCATCTCTCTCACTACATCTACATGAGCACTTCTGAGCACTTCTGAGAGAGCTG 7662  
 Qy 2463 HisIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2482  
 Db 7663 CATGCTCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7722  
 Qy 2483 TyMetIleuGlyIleuGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2502  
 Db 7723 TACGTGTGTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7782

Qy 2503 GluGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2522  
 Db 7783 CAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7842  
 Qy 2523 PheAlaGlyProValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2542  
 Db 7843 TTTGGGCGGCGCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7902  
 Qy 2543 ArgAlaSerCysAlaAlaGlnArgIleuIleuIleuIleuIleuIleuIleuIleuIleu 2562  
 Db 7903 AAGTTTCTGCTCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7962  
 Qy 2563 GlnProSerPheAlaValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2582  
 Db 7963 AGGACCGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8022  
 Qy 2583 ValAsnSerAspThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2602  
 Db 8023 GTGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8082  
 Qy 2603 ProPheIleuPheIleuSerTyValIleuIleuIleuIleuIleuIleuIleuIleu 2621  
 Db 8083 CCTTCTGCTCTCTTTCACATGCTGCTCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 8142  
 Qy 2622 LeuAlaCysSerArgIleuProSerProAspProAlaIleuIleuIleuIleuIleu 2640  
 Db 8143 GTCCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8202  
 Qy 2641 ThrSerSerTyArgCysProSerProTyAlaAspGly-----ArgIleuIleuIleu 2658  
 Db 8203 AGCGCTCTCTCACTGACGACACCACTTCGCTGAGCGGCGCTGACATGCTGCGACAGAC 8262  
 Qy 2659 TyGlyAspSerAlaGlySerIleuIleuIleuIleuIleuIleuIleuIleuIleu 2672  
 Db 8263 TTGGGAGAGTCCACCGCTCGCTGAGACGATCGTCAGAGGAGGAGGAGGAGGAGGAG 8322  
 Qy 2673 -----GlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2681  
 Db 8323 GCGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8382  
 Qy 2682 PheIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2701  
 Db 8383 AGAGCTGCAAGAT-----CCCCCTGGC----- 8406  
 Qy 2702 ProGlySerIleuPheIleuGlnGlyIleuIleuIleuIleuIleuIleuIleuIleu 2721  
 Db 8407 -----CACGATTCGATCTCAAT 8424  
 Qy 2722 SerAspIleuSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2741  
 Db 8425 AGCGAGCTGTCTCTG-----GATGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8481  
 Qy 2742 SerGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2761  
 Db 8482 AGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8517  
 Qy 2762 AspSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2780  
 Db 8518 GACCGCGGAGGAGG-----GCCGTCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8559  
 Qy 2781 GlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2800  
 Db 8560 GCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8619  
 Qy 2801 SerSerGlyAsnGlyAlaPro----- 2807  
 Db 8620 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8679  
 Qy 2808 -----GlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 2826  
 Db 8680 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8739



Db 1252 GAGGAGCGCGCCGAGTACAGCTCTGCTGGAGGCCACGACGAGGCGGCAATCCGGGC 1311  
 QY ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAsnAspAlaPro 397  
 Db 1312 CCGCTACGTGCCACGCGCCACCGTGTACTGAGGTGGAGGACAGAACCACTACCC 1371  
 QY GlnPheSerGlnIuArgGlyValValGlnValArgGluAspValThrProGlyAlaPro 417  
 Db 1372 CAGTTCAGGAGGAGAACTACGTGTCCAGGTGCCGAGGACGTGGGGCTCAACACGGCT 1431  
 QY ValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValAlaHisTyrSer 437  
 Db 1432 GTGCTCGAGTGCAGGCGACGAGCCGAGCCAGGCGCAACCGCGCACTTCACTACAC 1491  
 QY IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnIleuAsp 457  
 Db 1492 ATCTCAGCGGGAGACGTGGCCGCGGCTTCACTTCACTCGCTGAGCGGATCTGGAT 1551  
 QY ValValSerProLeuAspTyrGlnThrThrTyrGlyTyrThrLeuArgValArgAlaGln 477  
 Db 1552 GTGATCAACCCCTTGATTTGAGGATGTCAGAAATTACTCGTGAACATTAAGCCGACG 1611  
 QY AspGlyGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
 Db 1612 GATGGGGCGCGCGCCCGCTCATCTTCAAGGGGTGTCTGTGACGTGTGAGTCTGGAT 1671  
 QY IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
 Db 1672 GTCAACGACACGAGCTATCTTGTGACGAGCCCTTCCAGGCCACGCGTCTGGAGAT 1731  
 QY ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
 Db 1732 GTGCGCCCTGGGCTACCCCGTGTGACATTCAGCGGTGAGCGCGGACTGTGAGAGAAC 1791  
 QY AlaArgLeuGlnTyrArgLeuAla-----GlyValGlyHis--- 549  
 Db 1792 GCGCGCTGCACTATGCTGTGTGACACGCGCTTCACTTGTGGGGCGCGGACGCT 1851  
 QY 550 -----AspPheProPheThrIleAsnAsnGlyThrGly 560  
 Db 1852 GGGCTTAAGATCTGCGCCCAACCCCTGACTTCCCTTCCAGATTCACACAGCTCCGAT 1911  
 QY TrpIleSerValAlaAlaGluLeuAspArgGluGlnValAspPheTyrSerPheGlyVal 580  
 Db 1912 TGGATCAACGTGTGTCCGAGCTGTGACCGCGAGAGAGTGTGACATCAAGCTTCGGGTG 1971  
 QY GlnAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1972 GAGGCGGTGACCAACGCTGCGCCCAATGAGCTCTTCCACGCGTGTCCATCAACGGTG 2031  
 QY LeuAspValAsnAspAsnAsnProThrPheThrGlnProGlyTyrThrValArgLeuAsn 620  
 Db 2032 CTGGAGGTGAATGACACAGACCGCGTGTTCACGCAACCCCACTACAGACTTCCTCGAT 2091  
 QY GlnAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAlaHis 640  
 Db 2092 GAGGAGCGCGCTGTGGAGGACGCTGTGACCTCTCAGGCGCGCAACCGTGAACGCAAC 2151  
 QY SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 2152 AGGTGATTAACCTACAGCTCAGCGGGGACACCGGAAACCGCTTGTGACTCAACAGC 2211  
 QY GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrTyrLeuGlnValGln 680  
 Db 2212 CAGAGAGGGGGCGGCTCATCACTGTGGCTACTCTGGAATCAAGAGGAGGAGGAG 2271  
 QY TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 2272 TAGGTCTGGCGGTGACAGATCCGACGCGCAACGCTGTGCACTCGCATGTCCCAATC 2331  
 QY AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720

Db 2332 AACGTCACTGATGCCAACACCAAGGCTGTCTTTCAGAGCTCCCATTAACACTGAGT 2391  
 QY ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db 2392 GTCACTAGGACAGAGGCTGTGGGACCTTCACTGTCTACCTACGTAGGCCAACATAGAAC 2451  
 QY ThrGlyGluAsnAlaArgIleThrTyrPheMetGlnAspSerIleProGlnPheArgIle 760  
 Db 2452 ACAGAGAGGAATGCCCGCATCACTTACGATTCAGAGACCCCGTCCGCAAGTTCGGCAT 2511  
 QY AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db 2512 GACCCCGACAGTGGACCATGACATGACATGAGAGTGTGACTTACAGAACCAAGGTTCGCC 2571  
 QY TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnTyrSerAspThrThrTyr 800  
 Db 2572 TACAGCTGACATCATATAGGCCAGACACGCGATCCCGCAAAATCAGACACCAAC 2631  
 QY LeuGlnIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSerTyr 820  
 Db 2632 CTAGAGATCTCATCTCTGATGCCAATGACATGACATGCCAGTTCCTGTGGATTTCTAC 2691  
 QY GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2692 CAGGATCTCATCTTGAAGATGCTCCACCTTCAGACACAGCATCTCCAGGTCTTCGACG 2751  
 QY AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
 Db 2752 GACCGGACTCAGGTCCCAATGGCGGTCTGCTGTACCTTCCAGGCTGGGAGACACGCGC 2811  
 QY AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgArgLeuAsp 880  
 Db 2812 GATGGGAGCTTCTCATATGACACCCACGCGCGGTGATTCACACCAACGCGCGCTGAGC 2871  
 QY ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetPro--- 899  
 Db 2872 GCGGAAATGTGGCGGTGTACACTTGTGGCTGTGGCTGTGGATGGGGAGTCCACT 2931  
 QY ProAlaArgThrProMetGlnValThrValThrValLeuAspValAsnAspAsnProPro 919  
 Db 2932 CCCCTTACCGCGTGGTGAATTCAGGTGACATCTTGACATTAATGACATGCCGCC 2991  
 QY ValPheGlnGlnAspGlnPheAspValPheValGlnGluAsnSerProIleGlyLeuAla 939  
 Db 2992 ATGTTTGAGAGAGACGAACCTGAGTGTGTGTGAGAGAACACCCAGTGGGTGCGGTG 3051  
 QY ValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3052 GTGGCAAGATTCGTGTACGACCCCTGATGAAGCCCTTAATGCCAATCATGTATCAG 3111  
 QY IleValGlnGlyAsnIleProGlnValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3112 ATGTGGAAGGGGACATGCGGCAATTTCTTCCAGCTGTGACCTGTCAAGGGGACCTGGCT 3171  
 QY AlaLeuValAspLeuAspTyrGluAspArgProGlyTyrValLeuValIleGlnAlaThr 999  
 Db 3172 GCCATGTGTGAGCTGACTTTGAGGTCCGGCGGAGTATGTGTGTGTGTGTGTGTGTGTGT 3231  
 QY SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
 Db 3232 TCGGCTTCGCTGTGTGAGCGGACGAGTGCATCTTCTGTGTGTGTGTGTGTGTGTGTGT 3291  
 QY ProProValLeuGlyAsnPheGlnIleLeuPheAsnAsnTyrValThrAsnArgSerSer 1039  
 Db 3292 CCGCTGTGTGCGCCGACTTCCAGATCTTCAACACTATGTATCAACAAAGTCCAAAC 3351  
 QY SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
 Db 3352 AATTTCCTCCACCGGCGGTATGCGTCATCCCGGCGCAATGACCCGAGGTGTCAACAGC 3411  
 QY LeuThrTyrSerPheGlnArgGlyAsnGluLeuSerLeuValLeuLeuAsnAlaSerThr 1079  
 Db 3412 CTCAACTACACTTGT 3471

QY 1080 G1YGLIuLeuLYseuSerArgAlaLeuAspAsnAspArgProLeuGlnAlaIleMetSer 1099  
DB 3472 GGGAACTGAGCTCAGCCGCGACCTGGACAAACACCGCCGCTGGAGGCGCTCAATGAG 3531  
QY 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
DB 3532 GTGTCTGTCTGATGATGATCCACAGCTCAACCGCTTCTGACCGCTGCTGATCAATC 3591  
QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGlnAspMetSerProGlu 1139  
DB 3592 ATCAAGAGAGACATGCTGACCAACACATCACTGCGCTGGAGAACTGTCCCGAGAG 3651  
QY 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrIleuAlaThr 1159  
DB 3652 AGGTTCTGTCTGCGCTGCTGCGCCCTTCTGTGAGAGGGGTGCGCCGCTGTCTCAAC 3711  
QY 1160 ProProAspHisValValValPheAsnValGlnAspAspThrAspAlaProGlyGlyHis 1179  
DB 3712 ACCAAGAGACGATCTTCTGCTTCAACGTCACAGAACGACACCGACGTC---AGCTCCAAC 3768  
QY 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPhe 1199  
DB 3769 ATCTTGAAGTGAAGCTTCTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3822  
QY 1200 LeuProSerGluAspLeuGlnGlnArgLeuYrLeuAsnArgSerLeuLeuThrAlaIle 1219  
DB 3823 TTCCTGCTGAGAGACCTGACAGAGAGAGATCTAATTCGACGCTGCTGACACCAATC 3882  
QY 1220 SerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGluAsn 1239  
DB 3883 TCCAGCGCACGCGCGCTGCGCTTCCAGCAACATCTGCTGCGCGCGCGCGCGCGCGCGAGAAC 3942  
QY 1240 TyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSerSer 1259  
DB 3943 TACATGAAGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4002  
QY 1260 SerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProProGly 1279  
DB 4003 ACCGAGCTCTTCGCGCGCATCCACCCCAACAGCGCTGCGCTGCGCGCGCGCGCGCGCG 4062  
QY 1280 PheThrGlyAspTyrCysGluThrGlyValAspLeuCysTyrSerArgProCysGlyPro 1299  
DB 4063 TTCACCGCGGACTACTGCGACAGAGATCGACTCTGCTACTCCGACCCGCGCGCGCGCG 4122  
QY 1300 HisGlyArgCysArgSerArgGlnGlyGlyTyrThrCysLeuCysArgAspGlyTyrThr 1319  
DB 4123 AACGCGCGCTGCGCGACCGCGAGGGCGGCTACACTGCGAGGTGCTTGGAGGACTTCACT 4182  
QY 1330 G1YGLIuHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCysIlyAsn 1339  
DB 4183 GGAAGAGACTGTGAGGTGATCCCGCTCAGGCGCGCTGCGCAAGGGGTGTGCAAGAAC 4242  
QY 1340 G1YGLIYThrCysValAsnLeuValGlyGlyPheIlyCysAspCysProSerGlyAsp 1359  
DB 4243 GGGGCGACCTGTGTAACTGTCTACCTGCGCGCTTCACTGCTGTGTCTCTGCGGAG 4302  
QY 1360 PheGluYrProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIleThr 1379  
DB 4303 TATGAGAGGCTTACTGTGAGGTGACACACAGAGGCTTCCGCCCAAGCTTCTGTCAAC 4362  
QY 1380 PheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIlyGlu 1399  
DB 4363 TTCGGGGGCTGAGACAGGCTTCCACTTCAACATCTCCCTCAGCTTGTCCACTGAGAA 4422  
QY 1400 ArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyValHisAspPheValAlaLeu 1419  
DB 4423 AGGAACGGCTGTCTCTCAACACGCGCTTCAATGAGAACACAGCTTCACTCCCGCTG 4482  
QY 1420 GluValIleGlnGlnValGlnLeuThrPheSerAlaGlyGluSerThrThrVal 1439  
DB 4483 GAGATCGTGAACGACAGGTGACCTCACTTCTGTGAGGAGAACAAACAGACCGTG 4542

QY 1440 SerProPheValProGlyValSerAspGlyGlnThrPheIleThrValGlnLeuYrTyr 1459  
DB 4543 GCACGAAAGCTTCCAGTGTGTGTAGTACGGGCGGTGCACTGTGTGAGGTGACATAC 4602  
QY 1460 TyrAsnYrProLeuLeuGlyGlnThrGlyLeuProGlnGlyProSerGlnIlyVal 1479  
DB 4603 TACAACAGCCCAATATTGGCCACTGGGCGCTGCCCAATGGGCGCTCGGGGAAAAAGATG 4662  
QY 1480 AlaValValThrValAspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeu 1499  
DB 4663 GCCGTGTGACATGTGATGATGTGTGACAAACATGAGGTGTGTGCTTGGAAAGACATC 4722  
QY 1500 GlyAsnYrSerCysAlaAlaGlnGlyThrGlnGlyGlySerValSerLeuAspLeu 1519  
DB 4723 GGGAACTTACGCTGCGCTGCGCGACCTCAACCGGCTCCAAAGATCCTGTGATCTG 4782  
QY 1520 ThrGlyProLeuLeuLeuGlyValProAspLeuProGlySerPheProValArgMet 1539  
DB 4783 ACCGCGCTTACTCTCTGGGGGTGTCCCAACCTGCGCAAGACTTCCAGTGCACAAAC 4842  
QY 1540 ArgGlnPheValGlyCysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAla 1559  
DB 4843 CGGCAGTTCGTGGGTGTGATGCGGAACCTGTCACTGACGCGCAAAATGTGACATGCGC 4902  
QY 1560 AspPheIleAlaAsnAsnGlyThrValProGlyCysProAlaIlyValAsnValCysAsp 1579  
DB 4903 GGATTCATCGCCACAAATGGACCCCGGAGAGCTGCGCTGCGAGAACTTCTGCGAT 4962  
QY 1580 SerAsnThrCysHisAsnGlyYrThrCysValAsnGlnTyrAspAlaPheSerCysGlu 1599  
DB 4963 GGGAGCGGTGTGAATGAGAGGACCTGTGTCAACAGTGTGAATATGATCTGTGTAG 5022  
QY 1600 CysProLeuGlyPheGlyGlyIlySerCysAlaGlnGluMetAlaAsnProGlnHisPhe 1619  
DB 5023 TGTTCATCTCCATTTCCGGGGAGAACTGTGAGAACGCAATGCACTCCACCGCTCTTC 5082  
QY 1620 LeuGlySerSerLeuValAlaThrPheIleGlyLeuSerLeuProIleSerGlyProTyr 1639  
DB 5083 AGCGGTGAGAGCTGTGTCTTGGAGTGAACCTGAACATCACTCTGTGTGCTGCTGCTG 5142  
QY 1640 LeuSerLeuMetPheArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArg 1659  
DB 5143 CTGGGCTCATGTTCCGGACCCGGAAGGAGACAGCGTTGTGATGAGAGCCACAGTGT 5202  
QY 1660 GlyArgSerThrIleThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlyGly 1679  
DB 5203 GGGCCACACACTTTCGCTTCAATCTGTAACAACTACTCCAGTTGAGGTGCCAC 5262  
QY 1680 ThrGlyLeuGlnAlaSerSerLeuArgLeuGlnProGlyArgAlaAsnAspGlyAspTyr 1699  
DB 5263 GGGCCCTCCATGTGAGTCCGTGATGTCTGTCCGGTTGCCGTGACCGAGGAGTGTG 5322  
QY 1700 HisHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle 1716  
DB 5323 CACCACTGCTGATCGAGCTGAGATGATTTAAGAGACAGTGAATGAACACTGTGTC 5382  
QY 1717 ---LeuSerPheAspTyrGlyGlnGlnArgAlaGlnIlyAsnLeuGlyProArgLeuHis 1735  
DB 5383 ACATGACCTTGTGACTATGGAGTGAACCAAGAACAGAGATATCGGGGAGTCTTCC 5442  
QY 1736 GlyLeuHisLeuSerAsnIleThrValGlyGlyIleProGlyProAlaGlyValAla 1755  
DB 5443 GGGCTGACGGAAGAGAGCTGTGTGTGGAGGCCCTTGAAGACAGTCTCCGTGCC 5502  
QY 1756 ArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAspThrProGlyGlyValAsn 1775  
DB 5503 CGTGAATTCGAGGCTCAATGACAGGAGTGAAGTGGGGGAGGCCACCAAGTGTGCC 5562  
QY 1776 SerLeuAspProSerHisGlyGluSerIleAsnValGlnGlnGlyCysSerLeuProAsp 1795  
DB 5563 ACCTTGAACATGAACAAACGACTCAAGTCAAGGTGAAGACGCTGTGATGTGACGAC 5622  
QY 1796 ProCysAspSerAsnProCysProAlaAsnSerTyrCysSerAsnAspTyrAspSerTyr 1815



Db 5623 CCTGTACCTCGAGCCCTGTCTCCCAATAGCCGCTGCACAGCCGCTGGAGGACTAC 5682  
Qy 1816 SerCySerCyAspProGlyTyrTyrGlyAspAsnGlyThrAsnValCyAspLeuAsn 1835  
Db 5683 AGCTGGGCTGTGTAACAAGGGTACCTTGGAAATTAACGTGTGGATCTGTCCCTGAAC 5742  
Qy 1836 ProCyGluHisGlnSerValCyThrArgLeuProSerAlaProHisGlyTyrThrCys 1855  
Db 5743 CCTGGAGAAACATGGGGGCTGTGGCGCTCCCGGGCTCCCGAGGGCTACGTTGTC 5802  
Qy 1856 GlnCyProProAsnTyrLeuGlyProTyrCyGluThrArgTlleAspGlnProCyPro 1875  
Db 5803 GAGTGTGGGCGCCAGTCACTACGGGCGGTACTGTGAACAACACTCCACTTCGTGTC 5862  
Qy 1876 ArgGlyTyrTrpGlyHisProThrCyGlyProCyAsnCyAspValSerLeuGlyPhe 1895  
Db 5863 AGAGGCTGTGGGGAACCCCGTGTGACCTTGCACTGTGCGCTCCAGCAAGGCTTT 5922  
Qy 1896 AspProAspCyAsnLeuThrSerGlyGlnCyAsnLeuGlyAsnHisTyrArgPro 1915  
Db 5923 GATCCGAGCTGTATTAAGACCAACGGCAGTGCATGCAAGGAAATTAATAAGATC 5982  
Qy 1916 ProGlySerProThrCyLeuLeuCyAspCyTyrProThrGlySerLeuSerArgVal 1935  
Db 5983 CTAGCCAGGACACTGTCTGCTGCGCTGCACTGCTCCCACTGCCACAGCCCACT 6042  
Qy 1936 CyAspProGlnAspGlyGlnCyProCyLeuProGlyValTlleGlyArgGlnCyAsp 1955  
Db 6043 TGGCAATGGCCACCGGCGAGTGTGCTGCAAGCGGCGGCTACATCGGCGCCAGTCAAC 6102  
Qy 1956 ArgCyAspAsnProPheAlaGlnValThrThrAsnGlyCyGluValAsnTyrAspSer 1975  
Db 6103 CGCTGGACAAACCGTTTGCAGGTGCACACCGCTGGCTGTGAAGTATTAATGAG 6162  
Qy 1976 CyProArgAlaTlleGlnValGlyTleTyrTrpProArgThrArgPheGlyLeuProAla 1995  
Db 6163 TGTCCAAAGCATTTGAGCGGCGCATCTGTGTGCGCAAGCAAGATTCGCGGCGCT 6222  
Qy 1996 AlaAlaProCyProGlySerPheGlyThrAlaValArgHisCyAspArgHisTyr 2015  
Db 6223 GCGGTGCTAGCTCTTAAGGATTCGTTGAAATGCGGTGCGCACTGCGAGCGGGAGAA 6282  
Qy 2016 GlyTyrLeuProProAsnLeuPheAsnCyThrSerTlleThrPheSerGlyLeuGly 2035  
Db 6283 GGTGGGTGCGCCCAAGACTTTTAATCTATCAACATCTCTTCGAGCACTCAAGGCC 6342  
Qy 2036 PheAlaGlnArgLeuGlnArgGlnGlnSerGlyLeuAspSerGlyArgSerGlnGln 2055  
Db 6343 ATGAAATGAGAGCTGAGCCCGCAATGAGACCGAGGTGACCGCGCCGCTGCAGCTG 6402  
Qy 2056 AlaLeuLeuArgAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAspValTyr 2075  
Db 6403 GTGAGGCGGTGGCGAGTCTTACAGACACACGGGCAACGCTTTGGCAATGACGTGCG 6462  
Qy 2076 ValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGlnSerThrGlnArgGlyPhe 2095  
Db 6463 ACGGCTTACAGCTGTGGCGCAAGCTTCAAGACAGAGCTGGGCGAGGCTTCGAC 6522  
Qy 2096 LeuSerAlaThrGlnAspValHisPheThrGlnAsnLeuLeuArgValGlySerAla 2115  
Db 6523 CTGGACACCGACGAGACCGCGCACTTCAAGAGACCTCATCCACTCGGAGACGCGCTC 6582  
Qy 2116 LeuAspThrAlaAsnHisTyrGlnTyrGlnTyrGlnTyrGlnTyrGlnTyrAla 2135  
Db 6583 CTGGCCCAACGACACAGGGGCGGTGGAGACATTCAGCGGACGAGCGGCGGACGCA 6642  
Qy 2136 TyrLeuLeuGlnHisTyrGlnValTyrAlaSerAlaLeuAlaGlnAsnMetArgHis 2155  
Db 6643 CAGGTGCTGCGCGCTCGAGGGCTACTTACAGCAACGTGGACGCACTGGCGGCGGACG 6702  
Qy 2156 TyrLeuSerProPheThrTlleValThrProAsnTlleValTlleSerValValArgLeuAsp 2175

Db 6703 TACTGCGGCGCTTCGTATGCTACCGCCCAATGATTTCTGTGTGCATCTTTCAC 6762  
Qy 2176 LysGlyAsnPheAlaGlyValAlaLeuProArgTyrGlnValAlaLeuArgGlyGlnGlnPro 2195  
Db 6763 AGTTCAACTTACCGAGGCGAGGTCCGCGATTCGACACCATTCATGAAAGTTTCCC 6822  
Qy 2196 ProAspLeuGlnThrThrValTlleuProGlySerValPheArg-----Glu 2211  
Db 6823 ACGAGCTGAGTCTCCGCTCTTCCTCCAGCGCACTTCTTACAGCAACCTGAGAAAAA 6882  
Qy 2212 ThrProProValValArgProAlaGly-----ProGly 2222  
Db 6883 GAAAGCCCGCTGTGAGCGCGCTGGCGGAGAACACCCGCGAGACACCGCGCGGGG 6942  
Qy 2223 GlnAlaGlnGlnProGlnGlnTlleuAlaArgArgGlnArgTlleProGlnLeuSerGln 2242  
Db 6943 CTGGACACGAGAGGAGGAGCCCGATACAGACAGCGGAGGAGGAGACCTGATACGCTGGC 7002  
Qy 2243 GlyGlnAlaValAlaSerValTlleTyrArgThrTlleuAlaGlyLeuLeuProHis 2262  
Db 7003 CAGTTCGCGGTGCTGTGTATCATTTACGCACTGTGGGCGAGCTCCTGCGCGAGCGC 7062  
Qy 2263 TyrAspProAspLysArgSerLeuArgValProLysArgProTlleTlleAsnThrProVal 2282  
Db 7063 TACGACCCCGACCGTGCAGGCTCGGTGTGCTCAACGCGCCATTAATACCCGATG 7122  
Qy 2283 ValSerTlleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeuAspLysProVal 2302  
Db 7123 GTGACACGCTGTGTATACGAGGAGGAGCTCCGCTCCGAGACCCCTGAGAGGCGCTC 7182  
Qy 2303 ThrValGlnPheArgLeuLeuGlnThrGlnArgTyrTyrProLysCyAspAlaPheTrp 2322  
Db 7183 CTGTGTGAGTTGCGCTCTGTGAGGTGAGGAGGAGCAACAGCTGTGCTGCTGTCTGG 7242  
Qy 2323 AsnHisSerTlleLeuValSerGlyThrGlyTyrTyrProLysAlaArgGlyCyGluValVal 2342  
Db 7243 AACCACTCCCTGCGGTGTGTGAGACGAGAGGTGTGTGCGCGGCGTGCAGACTCTCG 7302  
Qy 2343 PheArgAsnGlnSerHisValSerCyGlnCyAsnHisMetThrSerPheAlaValLeu 2362  
Db 7303 TCCAGAGAACCGACACATGTGCTCGCAGTGCACACCAACAGCCAGCTTTCGGTGTCTC 7362  
Qy 2363 MetAspValSerArgArgGlnAsnGlyGlnTlleuProLeuLysThrLeuThrTyrVal 2382  
Db 7363 ATGATATCTTCCAGCGCGTGAACCGGAGGTCTCGCTGTGAATTTGTCACTATGCTC 7422  
Qy 2383 AlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheThrLeuLeuArg 2402  
Db 7423 GCTGTGCTTGTCACTGCGACGCTGCTGTGGCTGTGCTTCTCTGAGCTGTGCTGC 7482  
Qy 2403 TlleuArgSerAsnGlnHisGlyTlleArgArgAsnLeuThrAlaAlaLeuGlyLeuAla 2422  
Db 7483 ATGCTGCGCTTCAACTGACAGCATTCACAGACACTTGCAGGCGGTCTTCTCTCT 7542  
Qy 2423 GlnLeuValPheLeuLeuGlyTlleAsnGlnAlaAspLeuProPheAlaCyThrValTlle 2442  
Db 7543 CAGCTGTGTTGTGATTTGATGATCAACAGCAAGGAAACCCGTTCTGTGCAACATGTGT 7602  
Qy 2443 AlaTlleLeuLeuHisPheLeuTyrLeuCyThrPheSerTrpAlaLeuLeuGlnAlaLeu 2462  
Db 7603 GCAATCTCTCTCACTATCACTTACATGACACTTGTGCTGAGCCCTGTGAGAGCTTG 7662  
Qy 2463 HisLeuTyrArgAlaLeuThrGlnValArgAspValAsnThrGlyProMetArgPheTyr 2482  
Db 7663 CATGTCTACCGCATCTGACGAGGTGGCAACATGCAACGGGGCCATGTGGTTTAC 7722  
Qy 2483 TyrMetLeuGlyTyrGlyValProAlaPheTlleThrGlyLeuAlaValGlyLeuAspPro 2502  
Db 7723 TACGTGTGTGGGTGGGAGATCCCGGCATTTGCAAGAGCTGGCGGTGCTGAGACCCC 7782  
Qy 2503 GlnGlyTyrGlyAspProAspPheCyTrpLeuSerTlleTyrAspThrTlleTyrPse 2522  
Db 7783 CAGGCTACGGAGAACCCGACTTCTGTGCTGTGCTTCAAGACACCTGATTTGAGAC 7842





QY 32 pGlnValGlyProCysArgSerLeuGly---SerArgGlyAraGlySer----- 47  
 Db 59 -----CCCTGCGCGGAGATGAGGCTGCGAGCGCGCTGAGAGCGCGCTAACCC 108  
 QY 48 -----SerGlyAlaCysAlaPrometGlyTrp-----Le 57  
 Db 109 GCGCGGAGACCGCGCTTCGCGCTCCGCGCGCTGTAACTCAAGCGGTGAGCGCGCTTGC 168  
 QY 57 uCysProSerSerAlaSerAsnLeuTrpLeuTrpSer----- 70  
 Db 169 AGCGCCCGGAGCGCGGAGAGCTGAGAGCTGAGCGCGAGATGAGGCGGCTGAGAGAGCT 228  
 QY 71 -----ArgCysAraGAPAla-GlyTrpGlnLeuTr 80  
 Db 229 CGCGCGCTCTCGGCGCGGCGCGCGCTGCGCTCGCTCAAGTCCGCTTGTGTT----- 278  
 QY 80 hncGlyHisLeuValProHisAraGlyLeuAraGlyValTrpCysProGlnSerGlnAlaH 100  
 Db 279 --GCGCGCGAGTGCCTCC--GACGCGGCTGAGCGG-----CGCGCTGCGGCGCGC 324  
 QY 100 IsrLeuProLeuProProAlaProGln-GlyCysProTrpSerCysArgLeuLeuGlyIle 119  
 Db 325 AC-----GACCTTCCGCGCTGCGAGCGCGAGCCCGGCTCTGCGGAGAC 369  
 QY 120 GlyGlyHisLeuSerProGlnGlyLysLeuTrpLeuPro----- 132  
 Db 370 GGTGCGCGGCTGTC-----GGGAGCGCTCTGCTTCCGCTCCGCGGCGCTGCGGCGC 423  
 QY 133 GlnGlnHisProCysLeuLysAla----- 140  
 Db 424 GCGGAGCATTCGCGCGCTCCGAGCTCCGAGCACTTACCGCGCTGCGCGCGCGCGC 483  
 QY 141 ProArgLeuAraGlyCysGlnSer----- 147  
 Db 484 CCAGGCGCGCGCTGTCGCGCGCGCTCCCATCTGCTGCGCGCGGCGCTGCGCGCTG 543  
 QY 148 -----CysLysLeuAlaGlnAlaProGly---LeuAraGlyAgl----- 159  
 Db 544 CGTCTGCTGTGCGCTGCGCGCGCGCGCGCTGCGCGCTGCGGTGAGACTGCGCGTGAAG 603  
 QY 160 -----GluAraG 161  
 Db 604 GCGCGCACCGCGGAGACGCGCTCCGCGTCCCATCCCATCCGCGCGCTGCGCGCGAAC 663  
 QY 162 SerProGlnGlnSerLeuGlyAraGlyAraGlyAraGlyAraGlyAraGlyAraGly 179  
 Db 664 TTGCGCGGAGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGG 723  
 QY 180 -----GlnPheGlnProProSerTrpGlnAlaTrpValProGlnAraGlnProAlaGly 197  
 Db 724 AGCTGAGAGTTCCTCGATGCCCACTACAGGTGCGCTGTGTGAGAACCAACCGCGGCGC 783  
 QY 198 ThrProValAlaSerLeuAraGlyAlaLeuAraProAraGlyGlyAlaGlyAraGlyLeu 217  
 Db 784 ACCCTCATCTCCAGTGCAGCGCGCACTACACATGAGAGGCGAGAGAGCGCGTGAAGC 843  
 QY 218 TyrThrMetAspAlaLeuPheAraSerArgSerAsnGlnPhePheSerLeuAraProVal 237  
 Db 844 TATTATCATGAGAGGCGCTGTCGAGCGCTCCGCGGCTACTTCGGAATGCACTCTGCGC 903  
 QY 238 ThrGlyAlaValThrThrAlaGlnGluLeuAraPheGlyTrpLysSerThrHisValPhe 257  
 Db 904 AGCGGCGCGTGAAGCGAGCAAGCGTACGAGCGCGAGAACCAAGAGAGCGACGCTCTC 963  
 QY 258 ArgValThrAlaGlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIle 277  
 Db 964 AGGTGAAGCGGTGAAGCTACAGTACGCGCGCGCTCCGCGCACCACTACACTGCTGC 1023  
 QY 278 LeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnGlnLysLeuGlnSer 297  
 Db 1024 TTGGTCAAGAGACCAACAGCAACAGCGCGCTTCTGAGACAGTCCGAGTACCGCGAGCGC 1083  
 QY 298 LeuArgGlnAsnLeuValGlyTrpGlnValLeuThrValAraGlyAraThrAspGlyAraP 317

Db 1084 GTGCGGAGAGAACTGAGAGTGGCTACAGAGGTGCTGACCATTCGCGCGAGAGCGCGAAC 1143  
 QY 318 AlaProProAlaAsnAlaLeuTrpAraLeuLeuGlnGlySerGlyGlySerProSer 337  
 Db 1144 TCGCCATCAACGCCAATCTGCTTACCGCGGTGAGGAGGCGCGTGC----- 1191  
 QY 338 GlnValPheGlnLeuAraProAraSerGlyValIleAraGlyAraGlyProValAraPhe 357  
 Db 1192 GACGTCTTCAAGCTCAACGAGAGCTCTGCGCGGTGAGAACACCGCGCGGTGTGACCGG 1251  
 QY 358 GlnGlnValGlnSerTrpGlnLeuThrValGlnAlaSerAraGlnGlyAraPheProGly 377  
 Db 1252 GAGGAGCGCGCGAGTACAGCTCTGCTGAGGCGCAACGACGAGCGCGCATCCGGGC 1311  
 QY 378 ProArgSerThrThrAlaAlaValPheLeuSerValGlnAraAspAspAspAlaPro 397  
 Db 1312 CGCTCATGTGCCACCGCGCACCGGTATCATCGAGTGAAGAGACGAGAACCACTACCC 1371  
 QY 398 GlnPheSerGlnLysAraGlyTrpValValGlnValAraGlyAraPheProGlyAlaPro 417  
 Db 1372 CAGTTCAGCTGAGAGAACCTACGTGTCCAGGTGCCGAGAGAGTGGGCTCAACCGGCT 1431  
 QY 418 ValLeuAraGlyValThrAlaSerAraPheAraPheGlySerAsnAlaValAlaHisTrpSer 437  
 Db 1432 GTGCTGCAAGTGCAGGCGCACGAGCGGAGCACGAGCGCGCATTCATCACTACAGC 1491  
 QY 438 IleMetSerGlyAsnAlaAraGlyGlnPheTrpLeuAraPheAlaGlnTrpGlyAlaLeuAraP 457  
 Db 1492 ATCTCAAGCGGAGAGAGTGGCGCGCGCATTTCTACCTGCACTGCTGAGCGGAGATCTGAGT 1551  
 QY 458 ValValSerProLeuAraPheTrpGlnThrThrLysGlnTrpThrLeuAraValAraGln 477  
 Db 1552 GTGATCAACCCCTTGATTTCCAGAGATGTCAGAAATATCTGCTGAGACCTTAAGCCGAG 1611  
 QY 478 AspGlyGlyAraGlyProProLeuSerAraValSerGlyLeuValThrValGlnValLeuAraP 497  
 Db 1612 GATGGAGGCGCGCGCGCGCTCATCAATCTTCAAGGAGGTGTCTGTCAGAGTGTGAGT 1671  
 QY 498 IleAsnAraPheAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
 Db 1672 GTCAACGACCAAGAGCTATCTTGTGAGAGCGCGCTTCCAGAGCACAGGTGTGAGAGAT 1731  
 QY 518 ValProLeuGlyTrpLeuValLeuHisValGlnAlaLeuAraPheAraPheAraPheAraPhe 537  
 Db 1732 GTGCGCGTGCCTACCCCGGTGTGACATTCAGGCGGTGAGACGCGCATCTGAGAGAAC 1791  
 QY 538 AlaArgLeuGlnLysTrpArgLeuAla-----GlyValGlyHis--- 549  
 Db 1792 GCGCGGTGCACTATCGCTGTGTGAGACAGGCTTCCACTTCTGCGGCGGCGAGCGCT 1851  
 QY 550 -----AspPheProPheThrIleAsnAsnGlyTrpGly 560  
 Db 1852 GGGCTTAAGATTCCTGCGCGCGCACCGCTGACTTCCCTTCCAGATCAACAAGCTCGGCT 1911  
 QY 561 TrpIleSerValAlaAlaGlnLeuAraPheGlnGlnValAraPheTrpSerPheGlyVal 580  
 Db 1912 TGATCAACAGTGTGCGGAGCTGAGCCGCGAGAGGTGAGAGCACTACAGCTTCCGGGTG 1971  
 QY 581 GlnAlaAraGlyAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600  
 Db 1972 GAGGCGGTGAGACAGGCTGCGCGCGCGCATGAGCTTCCACAGGTGTCCATCAAGGAG 2031  
 QY 601 LeuAraPheAraAspAsnAraPheProThrPheThrGlnProGlnLysTrpValAraGlyLeuAra 620  
 Db 2032 CTGAGACGAGATGACCAAGACCGGTGTTCACGAGCGCACCTACAGCTTCTGCTGAT 2091  
 QY 621 GlnAraPheAlaAlaGlyThrSerValValThrValSerAlaAlaAraPheAraPheAraPhe 640  
 Db 2092 GAGGATGGCGCGGTGGAGAGAGGTGTGACCTGTGAGCGCGCGCGAGACCGGACCCCAAC 2151  
 QY 641 SerValIleThrTrpGlnIleThrSerGlyAraThrAraGlyAraPheSerIleThrSer 660

Db 2152 AGTGTGATTACTTACCACTGACAGGCGGCAACCCGGAACCGCTTGCACTCAGACG 2211  
Qy 661 GlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleuGln 680  
Db 2212 CAGAGAGGGGGGGGCTCATCAACCTGGGCTTACCTGGACTCAAGAGGAGGAGG 2271  
Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
Db 2272 TACGGCTGGCGGTGACACATCCGACGCGGACACGCTGCCACCTGCGCATCTCCAAATC 2331  
Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
Db 2332 AACGTCACTGATGCCAACAACCCAGGCTGTCTTCAAGCTCCCATTTACACAGTAAT 2391  
Qy 721 ValAsnGluAspArgProAlaGlyThrThrValIleLeuIleSerAlaThrAspGluAsp 740  
Db 2392 GTCAAGTAGAGACAGGCTGTGGGACCTGCATTTGCATCCCTGATGCCAAGATAGAGAC 2451  
Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
Db 2452 ACAGAGAAATGCGCGCATCACTACGTATTCAGAGACCCCGTCCGACGTTCCGCAAT 2511  
Qy 761 AspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
Db 2512 GACCCCGACAGTGGCACCATGTACACCATGATGAGCTGACTGATGAGAACAGGTCCGC 2571  
Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
Db 2572 TACACGCTGACCATCATGCGCCAGAGCAACGCGCATCCGCAAAATCAGACACCAACC 2631  
Qy 801 LeuGluIleLeuValAsnAspValAsnAspAlaProGlnPheLeuArgAspSerTyr 820  
Db 2632 CTAGAGATTCCTCATCTCGATGCCAATGACATGACACCCCAAGTTCCTGGGATTTCTAC 2691  
Qy 821 GlnGlySerValTyrGlyLeuAspValProPheThrSerValLeuGlnIleSerAlaThr 840  
Db 2692 CAGGCTTCATCTTGAGAGATGCTCAACCTCGACACGATCTCCAGGTCTCTCCACG 2751  
Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyGlyAspAspGly 860  
Db 2752 GACCCGGGACTCAGGTCCCAATGGCGTCTGTACACTTCAGAGGTGGGAGCAACGCGC 2811  
Qy 861 AspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgLeuAsp 880  
Db 2812 GATGGGAGCTTCTATCATCGAGCCACGTCGGGTGATTCGCAACGACGCGCGCTGAC 2871  
Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspIleGlyPhePro--- 899  
Db 2872 CGGAGGAATGGCGCTGATCAACCTTGCGGCTGTGGCTGTGGATCGGGGACGTCCACT 2931  
Qy 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspPro 919  
Db 2932 CCCCTTACCGCTCGGTAGAAATCCAGGTGACCATCTTGGAATATGACAAATGCCCCC 2991  
Qy 919 ValPheGluGlnAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 2992 AGTTTGAAGAGGAACTGAGCTGTTTGTGAGAGAACACACCAAGTGGGTGGGTG 3051  
Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3052 GTGGGAAAGATTCCGCTTACGACCTGATGAGGCCCTATGATCCCAATCATATGATGAG 3111  
Qy 960 IleValGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGlyLeuThr 979  
Db 3112 ATTTGGAGAGGGGACATCGGCAATTTCTCCAGCTGACCTGCTCAACGGGGACCTGGGT 3171  
Qy 980 AlaLeuValAspLeuAspTyrGluAspArgProGluTyrThrValLeuValIleGlnAlaThr 999  
Db 3172 GCATAGTGAAGCTGACTTGAAGTCCGGCGGAGTATGTGCTGTGTGAGGACCGACG 3231  
Qy 1000 SerAlaProLeuValSerArgAlaThrValHisValAlaGlyLeuLeuAspArgAspAsn 1019  
Db 3232 TCGGCTCGCGGTGAGCGGAGCCAGGTGACATCTTCTGTGTGACCAAGATGACAAAC 3291

Qy 1020 ProProValLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValThrAsnArgSer 1039  
Db 3292 CGGCTGTGTGCTGCCGACTTCAGATCTCTTCAACAATATATGATCAACAAGTCCAAC 3351  
Qy 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCCACCGGCGTATGCGGTGATCGGCGGCGCATGACCCGAGGTGTCAAGACG 3411  
Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerIleValLeuAsnAlaSerThr 1079  
Db 3412 CTCAACTACCTTGTGTGAGGAGGCAACGAGCTGCCCTGTGTGTGCTGAGACCCGACAG 3471  
Qy 1080 GlyGluLeuGlyLeuSerArgAlaLeuAspAsnAsnAspProLeuGluAlaIleMetSer 1099  
Db 3472 GCGCAACTGCAAGCTCAGCGCGAGCTGACACAAACCGCGCGCTGAGAGCGCTCATGAG 3531  
Qy 1100 ValLeuValSer---AspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThr 1118  
Db 3532 GTGTCTGTGTGAGATGGATGCATCCACGCGTCAAGCGCTTCTGACCCCTGCTGTACAC 3591  
Qy 1119 IleIleThrAspGluMetLeuThrHisSerIleThrLeuArgGluAspMetSerPro 1138  
Db 3592 ATCATCAAGACATGCTGACCAACAGCATCTGCTCCGCTGGAGAAACATGTTCCAG 3651  
Qy 1139 GluArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAla 1158  
Db 3652 GAGAAATTCCTGTCCCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3711  
Qy 1159 ThrProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGly 1178  
Db 3712 ACCACCAAGACAGAGTCTTCTGCTTCAACGTCCAGAACAGACACCGACGTC---AGCTCC 3768  
Qy 1179 HisIleLeuAsnValSerLeuSerValGlyGlnProProGlyGlyGlyProPro 1198  
Db 3768 AACATTCCTGAACGTGACCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3822  
Qy 1199 PheLeuProSerGluAspLeuGlnGluArgLeuTyrLeuAsnArgSerLeuLeuThrAla 1218  
Db 3823 TTCTTCCCGTGGAGGAGCTGACAGAGAGATACCTGATTCGATTCGAGCTGTACCAAC 3882  
Qy 1219 IleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCysGly 1238  
Db 3883 ATCTTCACGACGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3942  
Qy 1239 AsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSer 1258  
Db 3943 AACTCATGAATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4002  
Qy 1259 SerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysProPro 1278  
Db 4003 ACCACCGGTGCTTCCGCGCATCAACCCATCAACAGCGCGCTGCGCTGCGCTGCGCC 4062  
Qy 1279 GlyPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCysGly 1298  
Db 4063 GCGTTCACCGCGCATCTGCTGAGACGAGATCAACCTGCTACTCCGACCCGCGGCGC 4122  
Qy 1299 ProHisGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGlyTyr 1318  
Db 4123 GCCAAGCGCGCTGCGGACCGGAGGCGGAGGCGCTGACCTGCGAGTCTTGAAGATTC 4182  
Qy 1319 ThrGlyGluHisCysGluValSerAlaArgSerGlyArgCysThrProGlyValCysIle 1338  
Db 4183 ACTGAGAGGACTGTGAGGTGAGTCCGCTCAGGCGCTGTGTGCAACGGGATGTGCAAG 4242  
Qy 1339 AsnGlyGlyThrCysValAsnLeuValGlyGlyPheCysArgCysProSerGly 1358  
Db 4243 AACGGGGCACCTGCGTGAACCTGCTCATCGGCGGCTTCACTGCTGTGCTCCTGCGC 4302  
Qy 1359 AspPheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSerPheIle 1378  
Db 4303 GAGTATGAGAGGCGCTTACTGTGAGTGAACCAACGAGAGCTTCCGCGCCCACTCTTGTG 4362

QY 1379 ThPheArgGlyLeuAlaGlnInaGlnPheHisPheThrLeuAlaLeuSerPheAlaThrLys 1398  
 DB 4363 ACCCTCCGGGGCCCTGAGACAGCGCTCTCACCTTCACCTCCCTCAGTTGGCCACTGAG 4422  
 QY 1399 GlnAArgAArgGlyLeuLeuLeuLysArgGlyValArgPheAArgGlnLysSerPheValAla 1418  
 DB 4423 GAAAGGAACGGCTGCTCTCTCTACACGGCCGGCTTCATAGAAAGACAGACTTCATCGCC 4482  
 QY 1419 LeuGlyValIleGlnGlnGlnValAlaLeuThrPheSer-----AlaGlyGlySerThr 1436  
 DB 4483 CTGGAGATCCGAGACGAGGCTGACCTCCTCTCTGAGGTCGACGGAGACAA 4542  
 QY 1437 ThrThrValSerProPheValProGlyGlyValSerAArgGlyGlnThrValGln 1456  
 DB 4543 ACGACCTGGACCGAAGAGTTCACAGTGTGTGATGACGCGCGTGGCACTGTGTCAG 4602  
 QY 1457 LeuLysThrLysLeuLys-----ProLeu 1464  
 DB 4603 GTGCGAGTACTACAAAGGTAAATGGGCCCCCACTTCCCTGGCCCCCAGCCCAAT 4662  
 QY 1465 LeuGlyGlnThrGlyLeuProGlnGlyProSerGlyLysValAlaValAlaThrVal 1484  
 DB 4663 ATTGGCCACCTGGGCTGGCCCATGGCCGTCCGGGAAAAGATGGCCGTGTACAGTG 4722  
 QY 1485 AArgGlyCyAArgThrGlyValAlaLeuAArgPheGlySerValLeuGlyAArgLysSerCys 1504  
 DB 4723 GATGATGTGTACACAACTGGCTGTGCTGTGGAAAGACATCGGGAATCACTACGCTGC 4782  
 QY 1505 AlaAlaGlnGlyThrGlnGlnGlyLysSerLysSerLeuAArgPheThrGlyProLeuLeu 1524  
 DB 4783 GCTGCCAGAGGACTCAGACCGGCTCCAAAGATGCTGTGATCTGACCGGCTCTACTC 4842  
 QY 1525 LeuGlyGlyValProAArgLeuProGlyLysSerPheProValAlaGlnMetAArgGlnPheValGly 1544  
 DB 4843 CTGGGGGGGTGCCCCCACTGCCAGAGAATCTCCCACTGACAAACCGGAGTGTGTGGGC 4902  
 QY 1545 CyMetAArgAArgLeuGlnValAArgSerAArgHisIleAArgMetAlaAArgPheIleAlaAArg 1564  
 DB 4903 TGCATGGGAACCTGTGATGTCAGCGCAAAATGTGACATGGCGGATTCATCGCCAAAC 4962  
 QY 1565 AArgGlyThrValProGlyCyAProAlaLysLysValAArgValCyAArgSerAArgThrCysHis 1584  
 DB 4963 AATGGACACCGGGAAGGCTGGCTGCTCGAGAGAACTTTCGATGGGAGCGGCTGTACG 5022  
 QY 1585 AArgGlyGlyThrCysValAArgGlnIleAArgIleAArgAArgSerCysGlyCysProLeuGlyPhe 1604  
 DB 5023 AATGGAGGACCTGTGTCAACAGGTGGAATGTATCTGTGTAGTGTCCACTCCGATTC 5082  
 QY 1605 GlyGlyLysSerCysValAArgGlnIleMetAlaAArgProGlnHisPheLeuGlySerSerLeu 1624  
 DB 5083 GCGCGGAAGAACTGTAGCAAGGCTCTCAACCCCAAGCTTCAGCGGTGAGAGCGTTC 5142  
 QY 1625 ValAlaIleThrIleGlyLeuSerLeuProIleSerGlnProThrLysLeuSerLeuMetPhe 1644  
 DB 5143 GTGTCTCAGAGTACCTGAACATCATCTCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 5202  
 QY 1645 AArgThrAArgGlnAlaAArgGlyValLeuLeuGlnAlaIleThrAArgGlyAArgSerThrIle 1664  
 DB 5203 CGGACCGGGAAGGAGCAAGCGTGTGTATGAGGCGCAACAGTGTGGGCCCAACAGCTTC 5262  
 QY 1665 ThrLeuGlyLeuAArgGlyGlnIleValMetLeuSerValGlnGlyThrGlyLeuGlnAla 1684  
 DB 5263 CGCTCCAGATCTGAACAACTACTCTGATGAGGTGTCCACGGCCCTCCGATGTG 5322  
 QY 1685 SerSerLeuAArgLeuGlnProGlyAArgAlaAArgAArgLysAArgThrHisAlaGlnLeu 1704  
 DB 5323 GAGTCCGATGATGCTGTGCGGCTGTGCGGTCACGAGGAGAGTGTGCAACCTGTGTATC 5382  
 QY 1705 AlaLeu-----GlyAlaSerGlyGlyProGlyHisAlaIle-----LeuSerPheAArg 1720  
 DB 5383 GAGCTGAAGATGTTAAGAGAGACATGATGAAGACCTGTGTCAACATGACCTTGTGAC 5442  
 QY 1721 TyrGlyGlnGlnAArgAlaGlnGlyAArgLeuGlyProAArgLeuHisGlyLeuHisIleLeuSer 1740

DB 5443 TATGGATGACCAAGAAACAGGACATATCGGGGCGATCTTCCCGGCTGACGCTAAAG 5502  
 QY 1741 AArgIleThrValGlyGlyLysProGlyProAlaGlyGlyValAlaAArgGlyPheAArgGly 1760  
 DB 5503 AGCGTGTGTGAGAGGCGCTCTGAAGCAAGGTCTCGGTGGCCGTGTATTCGAGGC 5562  
 QY 1761 CysLeuGlnGlyValAArgValSerAArgThrProGlnGlyValAArgSerLeuAArgProSer 1780  
 DB 5563 TGCATCAGAGGATGAGATGGGGGGGAGCCCAACCAAGTGTGCACTGCAACCTGTGAACATGAAC 5622  
 QY 1781 HisGlyGlySerIleAArgValGlnGlnGlyCysSerLeuProAArgProCysAArgSerAArg 1800  
 DB 5623 AACGACCTAAGGTCAAGGTGAAGAGCGGCTGTGTATGTGAAGACCCCTGTACTTGAGC 5682  
 QY 1801 ProCysProAlaAArgSerLysThrPheAArgAArgThrAArgSerLysSerCysAArgCysAArg 1820  
 DB 5683 CCTGTCTCCCAATATGCGCTGTGCAACAGCGCTGGAGAGATCAAGCTGTGTGTAC 5742  
 QY 1821 ProGlyLysThrGlyAArgAArgCysThrAArgValCysAArgLeuAArgProCysGlnHisGln 1840  
 DB 5743 AAAGGCTACTTGAATTAATGTGTGATGCTGTCACTGAACCTGTGCAAGAACATG 5802  
 QY 1841 SerValCysThrAArgLysProSerAlaProHisGlyLysThrCysGlyLysProProAArg 1860  
 DB 5803 GGGGCTGTGTGCTGCCCGGCTCCCGAGGCTCAAGGCTGATGTCAGTGTGGGCCAGT 5862  
 QY 1861 TyrLeuGlyProLysCysGlnThrAArgIleAArgGlnProCysAArgProGlyLysThrGly 1880  
 DB 5863 CACTACGGGCGCTACTGTGAAGAAACAACTGCACCTTCCGTGCCAAGAGCTGTGTGGGG 5922  
 QY 1881 HisProThrCysGlyProCysAArgCysAArgValSerLysGlyPheAArgProAArgCysAArg 1900  
 DB 5923 AACCCGCTGTGTGACCTGTGCACTGTGTGCTGTGCAAGAAAGGCTGTGTATCCGCTGTATAT 5982  
 QY 1901 LysThrSerGlyGlnCysHisGlyCysLysGlnAArgHisLysAArgProProLysSerProThr 1920  
 DB 5983 AAGACCAAGGCGCAAGTGCATATGCAAGAAATTAACAAAGCTCTTACGCCAGACACC 6042  
 QY 1921 CysLeuLeuCyAArgCysAArgLysProThrGlySerLeuSerAArgValCysAArgProGlyAArg 1940  
 DB 6043 TGTCTCCCTGTGAGTCTGTCCCAATGGCTCCCAAGCGGCACTTGCAATGGCACCC 6102  
 QY 1941 GlyGlnCysProCysLysProGlyValIleGlyAArgGlnCysAArgAArgCysAArgAArgPro 1960  
 DB 6103 GGGCACTGTGTGCTGCAAGCGCGCTGCATCGCGCGCACTGTGACACCGCTGTGCAACACCG 6162  
 QY 1961 PheAlaGlnValThrThrAArgGlyCysGlyValAArgLysAArgSerCysAArgProAlaIle 1980  
 DB 6163 TTTCGCAAGGTCAACGCTGCTGTGAAGTATCTACATAGCTGTGTCCAAAGCATTT 6222  
 QY 1981 GlnAlaGlyLysThrProAArgThrAArgPheGlyLeuProAlaAlaAlaAArgCysAArgPro 2000  
 DB 6223 GAGCGCGGATCTGTGTGCGCAAGACAAAGTTCGGAACCGGCTGTGCGCATGCCCT 6282  
 QY 2001 LysGlySerPheGlyThrAlaValAArgHisCysAArgGlnHisAArgGlyLysLeuProPro 2020  
 DB 6283 AAGGAGTCCGTTGGAATCCGCTGCACTGACAGCGGGAAGAGGCTGTGCTGCCCA 6342  
 QY 2021 AArgLeuPheAArgCysThrSerIleThrPheSerGlnLeuLysGlyPheAlaGlnAArgLeu 2040  
 DB 6343 GAGCTTTTAATCTGTACACAACTCTCTGTGTGAACCTCAAGGCGCATGAATGAAGCTGT 6402  
 QY 2041 GlnAArgAArgLysSerLysLeuAArgSerGlyAArgSerGlnGlnLeuAlaLeuLeuAArg 2060  
 DB 6403 AGCGGATGAAGACGAGGTGGAGCGGCGCAGGCGCTGCACTGTGTGAGGGCGCTGCC 6462  
 QY 2061 AArgAlaThrGlnHisThrAlaGlyLysPheGlySerAArgValLysValAlaThrGlnLeu 2080  
 DB 6463 AGTGCTACACAGACACCGGCAAGCTCTTGTGCAAAAGATGTGTGCAACGGCTTACAGCTG 6522  
 QY 2081 AlaThrAArgLeuLeuAlaHisGlySerThrGlnAArgGlyPheGlyLysSerAArgThrGln 2100

Dh 6523 CTGGGCGACAGCTCTTCAGCAAGAGAGCTGGGAGGAGCTTCGACCTGGGAGCGACG 6582  
Qy 2101 AspValHisPheThrArgIleuAsnLeuArgValGlySerAlaLeuLeuAspThrAlaAsn 2120  
Db 6583 GAGCGGACATTTGACGAGGAGCGTCACTCGGCGAGCGGCTCCGCGGCCAGCGAC 6642  
Qy 2121 LysArgHisPheProGluLeuIleGlnGlnThrGluGlyGlyThrAlaPheLeuLeuGlnHis 2140  
Db 6643 AGGAGCGCGTGGGAGGAGTCCAGCGGAGCGAGCGGAGCGGAGCGGAGCGGAGCGG 6702  
Qy 2141 TyrGluAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisPheThrTyrLeuSerProPhe 2160  
Db 6703 CTGAGGCGGCTACTTCACAGCAAGTGGACCGGACGTCGCGCGGAGCGGAGCGGCTTC 6762  
Qy 2161 ThrTleValThrProAsnIleValIleSerValAlaArgLeuAspLysGlyAsnPheAla 2180  
Db 6763 GTCATCGTCACTCCGCAACATGTTCTGTGTGACATCTTTGACAAAGTTCACTTTACG 6822  
Qy 2181 GlyAlaLysLeuProArgTyrGluAlaLeuArgGlyGlyGlnProProAspLeuGluThr 2200  
Db 6823 GAGGCGAGGCTCCGCGATTCGACACCATCATGAAAGTTCCCGAGGAGCTGGAGTCC 6882  
Qy 2201 ThrValIleLeuProGluSerValPheArg-----GluThrProProValVal 2216  
Db 6883 TCCGTCTCTCCACGCGACCTTCTCAGACCACTGAGAGAAAGAGGCGCCCTGCTG 6942  
Qy 2217 ArgProAlaGly-----ProGlyGluAlaGlnGluPro 2227  
Db 6943 AGGCGCGCTGGCGGAGAACCAACCGCGAGACACAGCGCGCGCTGGCGACCGAGAG 7002  
Qy 2228 GluGluLeuAlaArgArgGlnAlaGlnHisPheGluLeuSerGlnGlyGluAlaValAla 2247  
Db 7003 GAGGCGCGCATCGACGAGCGGAGCGACCCCTGATGACCGTGGCGAGTTCGCGCT 7062  
Qy 2248 SerValIleIleTyrArgThrLeuAlaGlyLeuLeuProHisAsnTyrAspProAspLys 2267  
Db 7063 CTGCTATCATTTTACCGCACCTTCGGGAGAGCTCTGCGCGGCTACAGACCGCGAG 7122  
Qy 2268 ArgSerLeuArgValProLysArgProIleIleAsnThrProValValSerIleSerVal 2287  
Db 7123 CGGAGCTCCGCGTCTCAACCGGCCATCATTAATACCCGATGTGAGCAACGCGGTG 7182  
Qy 2288 HisAspAspGluGluLeuLeuProArgAlaLeuAspLysProValThrValGlnPheArg 2307  
Db 7183 TACAGGAGAGGAGCTCCGCTCCGAGACCCCTGAGAGGCGCTCTGCGTGGAGTTCG 7242  
Qy 2308 LeuLeuGluThrGluArgThrLysProIleCysValPheThrAsnHisSerIleLeu 2327  
Db 7243 CTGCTGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7302  
Qy 2328 ValSerGlyThrGlyGlyTyrSerAlaArgGlyCysGluValValPheArgAsnGluSer 2347  
Db 7303 GTTGGTGGAGCGGAGGAGGTGTGCTGCGGCGCTGGAGGCTCTGTCAGAGACGGA 7362  
Qy 2348 HisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArg 2367  
Db 7363 CATGTGCTGCGCAGTGCAGGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7422  
Qy 2368 ArgGluAsnGlyGlyIleLeuProLeuLysThrLeuThrTyrValAlaLeuGlyValThr 2387  
Db 7423 CCGAGAGACGGGAGGAGTCTGCTCTGAGATGTGCACTATGCGGCTGCTGTGTA 7482  
Qy 2388 LeuAlaAlaLeuLeuThrPhePheLeuThrLeuLeuArgGluLeuArgSerAsn 2407  
Db 7483 CTGGGAGCGCTGTGGCTTCGCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7542  
Qy 2408 GlnHisGlyTleArgArgAsnLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu 2427  
Db 7543 CTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7602  
Qy 2428 LeuGlyTleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis 2447  
Db 7603 ATTGGAGTACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7662

Qy 2448 PheLeuTyrLeuCysThrPheSerTrpAlaLeuLeuGluAlaLeuHisIleuTyrArgAla 2467  
Db 7663 TACATTTACATGAGGAGCTTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7722  
Qy 2468 LeuThrGluValArgAspValAsnThrGlyProMetLysPheTyrTyrMetLeuGlyTyr 2487  
Db 7723 CTGACGAGGAGTGGAGCATGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7782  
Qy 2488 GlyValProAlaPheLeuThrGlyLeuAlaValGlyLeuAspProGluIleTyrGlyAsn 2507  
Db 7783 GGATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7842  
Qy 2508 ProAspPheCysThrLeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProVal 2527  
Db 7843 CCGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7902  
Qy 2528 AlaPheAlaValSerMetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAla 2547  
Db 7903 GAGCTGTAT 7962  
Qy 2548 AlaGlnArgGlnGlyPheGluLysGlyProValSerGlyLeuGlnProSerPheAla 2567  
Db 7963 AAGAGCACCATTTAT 8022  
Qy 2568 ValLeuLeuLeuLeuSerAlaThrTrpLeuAlaLeuLeuSerValAsnSerAspThr 2587  
Db 8023 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8082  
Qy 2588 LeuLeuPheHisTyrLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
Db 8083 CTGAGCTTCACTACTCTCTCCGCACTTCACGCGGCTTACAGGCGCCCTTCCTCC 8142  
Qy 2608 SerTyrValValLeuSerLysGluValArgValAlaLeuLys---LeuAlaCysSerArg 2626  
Db 8143 TTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8202  
Qy 2627 LysProSerProAspProAlaLeuThrThrLysSerThr---LeuThrSerSerTyrAsn 2645  
Db 8203 AAGCTGACCTGAGAGGAGTCCGCGACCACTGAGAGGAGGAGGAGGAGGAGGAGGAG 8262  
Qy 2646 CysProSerProTyrAlaAspGly-----ArgLeuTyrGlnProTyrGlyAspSerAla 2663  
Db 8263 TCCACACCACTTGTGTGAGCGGCGCTGACATGCTGCGCACAGACTTGGGCGAGTCC 8322  
Qy 2664 GlySerLeuHisSerThrSerArgSer----- 2672  
Db 8323 GCTGCTGAGACGATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 8382  
Qy 2673 -----GlyLysSerGlnProSerTyrIleProPheLeuLeuArgGlu 2686  
Db 8383 CTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8442  
Qy 2687 GluSerAlaLeuAsnProGlyGlnGlyProProGlyLysAspProGlySerLeuPhe 2706  
Db 8443 -----CCCCCTGGC----- 8451  
Qy 2707 LeuGluGluGlnAspGlnGlnHisAspProAspThrAspSerAspLeuSerLeu 2726  
Db 8452 -----CAAGATTCAGCTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8484  
Qy 2727 GluAspAspGlnSerGlySerTyrAlaSerThrHisSerSerAspSerGluGluGlu 2746  
Db 8485 ---GATGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8541  
Qy 2747 GluGluGluGluGluGluAlaAlaPheProGlyGlnGlnGlyTyrAspSerLeuLeuGly 2766  
Db 8542 GTGGAGGCTGAGGAGAAA-----TGGAGCCCGGCGAGGAGG 8577  
Qy 2767 ProGlyAlaGluArgLeuProLeuHisSerThrProLys---AspGlyGlyProGlyPro 2785  
Db 8578 -----GCCGTCAAGACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8619



Db 904 AGCGGCGCGGTAGACAGCGCTACGACCGCGAGACCAAGACGACGACCTCTC 963  
 Qy 258 ArgValThrAlaGlnAspHisGlyMetProArgSerAlaLeuAlaThrLeuThrIle 277  
 Db 964 AGGGTGAAGCCGTGACCTACAGTACGCCGCCGCTCGGACCACTACATCACTGTG 1023  
 Qy 278 LeuValThrAspThrAsnAspHisAspProValPheGlnGlnGlnValTyrIleGlySer 297  
 Db 1024 TTGGTCAAGACACCAACGACCAAGCCCGGTCTTCAGACAGTGGAGTACCGGACGGC 1083  
 Qy 298 LeuArgGluAsnLeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAsp 317  
 Db 1084 GTGCGGAGAACCTGAGGTGGCTACAGAGTGTGACCACTCCGCGACCAACCGCAC 1143  
 Qy 318 AlaProProAsnAlaAsnIleLeuTyrArgLeuLeuGlnGlySerGlyIleSerProSer 337  
 Db 1144 TCGCCCAACACGACCACTTGGCTTACCCCGTGTGGGGGCGCGTGG----- 1191  
 Qy 338 GluValPheGluThrAspProArgSerGlyValIleArgThrArgGlyProValAspArg 357  
 Db 1192 GAGGTCTTCAGCTCAACGACGACCTCTGGCGTGTGAGACACAGGCGGTGTGACCGG 1251  
 Qy 358 GluGluValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGly 377  
 Db 1252 GAGGAGCGCGCGAGTACAGCTCTGTGTGAGGCGAACGACAGGCGCGCATCCGGGC 1311  
 Qy 378 ProArgSerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaPro 397  
 Db 1312 CCGCTCAATGCGACCGGCCACCTGTACATCAGAGTGTGAGACAGAACACATCACTACC 1371  
 Qy 398 GlnPheSerGluValArgTyrValValGlnValArgGluAspValThrProGlyAlaPro 417  
 Db 1372 CAGTTCACGCGAGCAAGACCTACGTGTCCAGGTGCCGAGGACGTGGGCTCAACACGCT 1431  
 Qy 418 ValLeuArgValThrAlaSerAspArgAspGlySerAsnAlaValValHisTyrTyr 437  
 Db 1432 GTGTGCGAGGTGACGACCGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 1491  
 Qy 438 IleMetSerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyValLeuAsp 457  
 Db 1492 ATCTCAGCGGGAGCGTGGCGCGGCAAGTTCATCCTGCACTCGGACGCGGATCTGGAT 1551  
 Qy 458 ValValSerProLeuAspTyrGlnThrTyrIleGlyTyrThrLeuArgValArgAlaGln 477  
 Db 1552 GTGATCAACCCCTTGATTTGAGGATGTCCAGAAATACCTCGTGAAGATTAAGGCCAG 1611  
 Qy 478 AspGlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAsp 497  
 Db 1612 GATGGGGGCGCGCCCGCTCATCAATCTTCAGGGGTGTGTGTGTGTGTGTGTGTGTGT 1671  
 Qy 498 IleAsnAspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGlnSer 517  
 Db 1672 GTCAACGACCAAGACCTTCTTCTGTGAGACGCCCTTCAGGACCGGTGTGTGTGTGTGT 1731  
 Qy 518 ValProLeuGlyTyrLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsn 537  
 Db 1732 GTGCGCGGTGGCTACCGCGTGTGTGACATTCAGCGGTGTGACGCGGATCTTGAGAGAAC 1791  
 Qy 538 AlaArgLeuGlnTyrArgLeuAla-----GlyValGlyHis--- 549  
 Db 1792 GCCCGGCTGACCTATCGCTGTGTGTGACACGCGCTCACCTTTCTGGGGGCGCGCGCT 1851  
 Qy 550 -----AspPheProPheThrIleAsnAsnGlyTyrGly 560  
 Db 1852 GGGCTTAAGAAATCTGCCCCACCCCTGACTTCCTTCAGATCAACAACAGCTCCGGT 1911  
 Qy 561 TrpIleSerValAlaAlaGluLeuAspArgGluValAspPheTyrSerPheGlyVal 580  
 Db 1912 TGGATCAACAGT 1971  
 Qy 581 GluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValSerValThrVal 600

Db 1972 GAGCGGTGTGACACAGGCTCGCCGCCCAATGAGCTCTCCACAGCGTGTCTATCAAGGTG 2031  
 Qy 601 LeuAspValIleAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeuAsn 620  
 Db 2032 CTGACGCTGAATGACCAACGACCCGCTGTTCACGAGGCCCACTACAGACTTGTGTGTAT 2091  
 Qy 621 GluAspAlaAlaValGlyThrSerValThrValSerAlaValAspArgAspAlaHis 640  
 Db 2092 GAGATGTGGCGCGTGGGGAGGACGCTGTGTGACCTTCAGGCGCCCGGACCGTGTGACGCAAC 2151  
 Qy 641 SerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThrSer 660  
 Db 2152 AGTGTGATTAACCTTACAGCTCAACGAGCGCAACCCCGGAACCCCTTGTGACCTAGCGAC 2211  
 Qy 661 GlnSerGlyGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrIleValLeuGluArgGln 680  
 Db 2212 CAGAGAGGGGCGGCTCATACCTTCAGCTTGTGACTTGTGACTGTGACTGTGACTGTGACT 2271  
 Qy 681 TyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleValVal 700  
 Db 2272 TACGTGTGGCGGTGTGACGATCCGACGCGACAGGTGTGACACTGGCGCATGTCTATC 2331  
 Qy 701 AsnValThrAspAlaAsnThrHisArgProValPheGlnSerSerHisTyrThrValAsn 720  
 Db 2332 AACGTCACTGATGCGAACCCACAGCGCTGTCTTTCAGAGCTCCATTCACAGTGAAT 2391  
 Qy 721 ValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrAspGluAsp 740  
 Db 2392 GTCACTGAGAGACGCTGTGTGGGACCTTCATCTCACTTCACTTCACTTCACTTCACTTCA 2451  
 Qy 741 ThrGlyGluAsnAlaArgIleThrTyrPheMetGluAspSerIleProGlnPheArgIle 760  
 Db 2452 ACAAGAGAGATGCGCCGATCACTCACTGATTCAGAGCCCGGCGCGCATTCGCAAT 2511  
 Qy 761 AspAlaAspThrGlyValAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnValSer 780  
 Db 2512 GACCCGACAGTGTGACATGATACACATGATGAGTGTGACATGATGAGAACAGAGTGTGCG 2571  
 Qy 781 TyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThrTyr 800  
 Db 2572 TACACGCTGACACATCAGGCGCCAGACAGAGCATCCCGAGAAATACAGACCAACAC 2631  
 Qy 801 LeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuAspSerTyr 820  
 Db 2632 CTAAAGATCTCATCTCCGATGCGCAATGACATCAACCCAGTTCGTGTGGATTTCTAC 2691  
 Qy 821 GlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAlaThr 840  
 Db 2692 CAGGTTCCATCTTTGAGATGTCTCACCTTCAGACGACATCTTCAGAGTCTGTGACAG 2751  
 Qy 841 AspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnGlyValAspAspGly 860  
 Db 2752 GACCGGACCTCAGTCCCATGTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2811  
 Qy 861 AspGlyAspPheIleValGluSerThrSerGlyIleValArgThrLeuAspArgLeuAsp 880  
 Db 2812 GATGGGACCTTTCATCACTGAGGCCACGTCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2871  
 Qy 881 ArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlySerPro--- 899  
 Db 2872 CCGAGAGATGTGGCGGTGTACACCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2931  
 Qy 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAsnProPro 919  
 Db 2932 CCGCTTACCGCGTGTGAGAAATCCAGGTATCACTTTGAGCATTAATGACAAATGCGCCC 2991  
 Qy 920 ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
 Db 2992 ATGTTTGAAGAGACGAACTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3051  
 Qy 940 ValAlaArgValThrAlaThrAspProAspGlnGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3052 GTGGCAAAAGATTCGTGTCAACAGCCCTGATGAAGCCCTTAATGCCAGATCATATATCAG 3111



QY 960 ILVAlGluGlyAsnIleProGluValPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
Db 3112 ATGTGGAGAGGAGCATGGCGCATTTCTTCAGGTGAGCATGGCTCAACGGGGAGCTGGCT 3171  
QY 980 AlaleuValAspLeuAspIleGlyAspArgProGluValLeuValIleGlnIleThr 999  
Db 3172 GCCATGTGAGCTGAGCTTGTGAGTCCGGCGAGATGATGTGGTGGTGGCGGACG 3231  
QY 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3232 TCGGCTCCGTGGTGGACCGAGCCAGTGCATCTTCTCGTGGACCAAGATGACAC 3291  
QY 1020 ProProValLeuGlyAsnPhenGlyIleLeuPheAsnAspIleValThrAsnArgSer 1039  
Db 3292 CCGCGTGTGCGCCGACCTTCAGATCTCTTCAACAATATGTCACCAAGTCCAC 3351  
QY 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAspIleSerAspSer 1059  
Db 3352 AGTTTCCCAACCGCGGTGATCGGTGCATCCGCGCATGACCCGACGTGTGACAGC 3411  
QY 1060 LeuThrTyzSerPheGluArgGlyAsnGlyLeuSerValLeuLeuAsnAlaSerThr 1079  
Db 3412 CTCACTACACCTTCGTGGAGGCGACAGCTGGCTGTGGTGGACCCCGCACG 3471  
QY 1080 GlyGlyLeuLeuValSerArgAlaLeuAspAsnAspArgProLeuGlyAlaIleMetSer 1099  
Db 3472 GCGGAACGTGAGCTGAGCGCGACCTGACACACCGCGCGTGGAGGCGCTCATGAG 3531  
QY 1100 ValLeuValSer---AspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThr 1118  
Db 3532 GTGTCTGTGTCTGACAGTGCATCCACAGCGTCAACCGCTTCTGCAACCTCGCTGCACC 3591  
QY 1119 IleIleThrAspGlnMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerPro 1138  
Db 3592 ATCATACGAGCAACATGCTGACCAACAGCATCTGTCCGCTGGAGAACATGTCCAG 3651  
QY 1139 GlyArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaIleThrLeuAla 1158  
Db 3652 GAAAGATTCCTGTCCCGCTGCTGGCCCTCTTGTGAGGGGGTGGCGCGTGTCTTC 3711  
QY 1159 ThrProProAspHisValValPheAsnValGlnArgAspThrAspAlaProGlyGly 1178  
Db 3712 ACCACCAAGAGCAACGTCTTGTCTTCAACGTCCAAAGACCAACCAACGTC 3768  
QY 1179 HisIleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyProPro 1198  
Db 3769 AAGATCTGAACGTGACCTTCTGGCGCTGCTGCGCGCGCTCGCGCGC-----CAG 3822  
QY 1199 PheLeuProSerGlyAspLeuGlnGlyArgLeuTyLeuAsnArgSerLeuLeuThrAla 1218  
Db 3823 TTCTCCCGTCCGAGGAGCTGACAGAGCAATCTGATCTGAGAGCGTGTGACAC 3882  
QY 1219 IleSerAlaGlnArgValIleProPheAspAsnIleCysLeuArgGluProCysGlu 1238  
Db 3883 ATGTCCACGAGGCGGTGCTGCTTGCAGAACCACTGTGCGCGAGCGCTTGGAG 3942  
QY 1239 AsnTyzMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAlaSer 1258  
Db 3943 AACTTACATGAAGTGGTGTGCTGCTTGCATTCGACAGCTCCGCGCTTCTCACTTC 4002  
QY 1259 SerSerValLeuPheArgProIleHisArgValGlyLeuArgCysArgCysProPro 1278  
Db 4003 ACCACCGTCTCTCCGCGCCATCCACCCATCAACGCGCTGCGCTCGCGCGC 4062  
QY 1279 GlyPheThrGlyAspTyzCysGlnThrGlnValAspLeuCysTyzSerArgProCysGly 1298  
Db 4063 GGGCTTCAACCGGACATCTGAGACGAGATGACCTTGTGCTACCTCGACCGGTGGG 4122  
QY 1299 ProHisGlyArgCysArgSerArgGlyGlyTyzThrCysLeuCysArgAspGlyTyz 1318  
Db 4123 GCCAAGCGCGCTGCGCGACCGCGAGGGGCTACACCTGCGAGTGTCTTCAGAGACTTC 4182

QY 1319 ThrGlyGlnHisCysGlyValSerAlaArgSerGlyArgCysThrProGlyValCysIys 1338  
Db 4183 ACTGAGAGACATGTGAGGTGATGCCGCTGACGGCGCTGTGGCCAAAGGGGTGGCAG 4242  
QY 1339 AsnGlyGlyThrCysValAsnLeuLeuValGlyGlyPheIysCysAspCysProSerGly 1358  
Db 4243 AACGGGGGACCTGTGGTGAACCTGTCAACGCGCGCTTCACTGCTGTGTCTCTGCGC 4302  
QY 1359 AspPheGluValProTyzCysGlnValThrThrArgSerPheProAlaHisSerPheIle 1378  
Db 4303 GAGTATGAGAGGCGCTTACTGTAGAGTACACAGAGCTTCCCGCCAGTCTTCTGTC 4362  
QY 1379 ThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAlaThrIys 1398  
Db 4363 ACCTTCCGGGCGCTGAGACAGCGCTTCCATTCACATCTCCCTCAAGTTCCACTCAG 4422  
QY 1399 GlyArgAspGlyLeuLeuLeuTyzAsnGlyArgPheAsnGlnIysHisAspPheValAla 1418  
Db 4423 GAAAGGAGCGGCTGTCTTCTTCAACGCGCTTCAATGAGAGCAACATTCATCGCC 4482  
QY 1419 LeuGluValIleGlnGluGlnValGlnLeuThrPheSer-----AlaGlyGluSerThr 1436  
Db 4483 CTGAGATGTGTGAGAGAGAGGTGACGTCACTTCTGTGACGTGCGAGCGAGCAACA 4542  
QY 1437 ThrThrValSerProPheValProGlyGlyValSerAspGlyGlnThrPheIleThrValGln 1456  
Db 4543 ACGACCGTGGACCGAAGGTTCCAGTGTGTGAGTGAAGGCGGTGGGCACTGTGGCAG 4602  
QY 1457 LeuIysTyzTyzAsnIys-----ProLeu 1464  
Db 4603 GTCGAGTACTTACAAAGAGTAAGTGGCCCAACATTCCTCCGCGCCCAAGCCCAAT 4662  
QY 1465 LeuGlyGlnThrGlyLeuProGlnGlyProSerGlnGlnIysValAlaValThrVal 1484  
Db 4663 ATTTGGCACTGGGCGCTGCGCCATAGGCGCTCCGGGAAAGATGGCGCTGTGACAGT 4722  
QY 1485 AspGlyCysAspThrGlyValAlaLeuArgPheGlySerValLeuGlyAsnTyzSerCys 1504  
Db 4723 GATGATTTGACACAACCATGCTGTGCGCTTGTGAAGAGCATGGGGAATCAACGCTGC 4782  
QY 1505 AlaIaGlnGlyThrGlnGlyIysSerIysSerLeuAspLeuThrGlyProLeuLeu 1524  
Db 4783 GCTGCCAGGCACTGACCGGCTTCAGAAAGTCCGTGAATCTGACCGGCGCTTACTTC 4842  
QY 1525 LeuGlyGlyValProAspLeuProGlyLeuSerPheProValArgMetArgGlnPheValGly 1544  
Db 4843 CTGGGGGGTGTCCCAACCTGCAAGAGACTTCCAGTGCACAAACCGGAGTTCCGTGGC 4902  
QY 1545 CysMetArgAsnLeuGlnValAspSerArgHisIleAspMetAlaAspPheIleAlaAsn 1564  
Db 4903 TGCATGCCGGAACCTGTCAGTCAACCGCAAAATGTGGCATGTGCGGATTCATCCCAAC 4962  
QY 1565 AsnGlyThrValProGlyCysProAlaIysValAsnValCysAspSerAsnThrCysHis 1584  
Db 4963 AATGGACCCGGAAAGGCTGCGCTGCGAGAACTTGTGAGTGGAGGCGGTGTCCAG 5022  
QY 1585 AsnGlyGlyThrCysValAsnGlnThrAspAlaPheSerCysGlyCysProLeuGlyPhe 1604  
Db 5023 AATGAGGCACTGTGTCAACAGTGAATATGTATCTGTGTGAGTGTCACTCCGATTC 5082  
QY 1605 GlyGlyIysSerCysAlaGlnIleMetAlaAspProGlnHisPheLeuGlySerSerLeu 1624  
Db 5083 GCGGGGAAGAACTGTGAGAACCACTGCTCAACCCCAAGCTTTCAGCGGTGAGAGCGTC 5142  
QY 1625 ValAlaThrPheIleGlyLeuSerLeuProIleSerGlnProTyzTyzLeuSerLeuMetPhe 1644  
Db 5143 GTGTCTGAGAGTCACTGAACATCACTCTTGTGCGCTGTGACTGTGGGCTTCATGTTTC 5202  
QY 1645 ArgThrArgGlnAlaAspGlyValLeuLeuGlnAlaIleThrArgGlyArgSerThrIle 1664  
Db 5203 CGGACCGGAGAGAGACAGCGCTTCTGATGAGGCGCAACAGTGTGTGGCGCCACGACTTC 5262  
QY 1665 ThrLeuGlnLeuArgGlnGlyHisValMetLeuSerValGlnGlyThrGlyLeuGlnAla 1684

Db 5263 CGCCTCCAGATCCCTGAACAACACTACTCCAGTTTAGAGGTGTCCCAAGGCCCTCCGATGTG 5322  
Qy 1685 SerSerLeuAlaGluInPProGlyValArgAlaAsnAspGlyAspTrpHisAlaGluLeu 1704  
Db 5323 GAGTCCTGGATGCTGTCCGGGTGGCGGTGACCCGACGGGAGATGGACCACTGTGTATC 5382  
Qy 1705 AlaLeu-----GlyAlaSerGlyProGlyHisAlaIle--LeuSerPheAsp 1720  
Db 5383 GAGCTGAAGAAATGTTAAGAGAGACAGTAGAATGAAGCACTGTGTACATGACTTGGAC 5442  
Qy 1721 TyrGlyGlnGlnArgAlaGluGlyValAsnLeuGlyProArgLeuHisGlyLeuHisIleuSer 1740  
Db 5443 TATGGAGATGAGACAGAAACAAGCAGATATCCGGGGCATGCTTCCGGGGCTGAACGGTAAG 5502  
Qy 1741 AsnIleThrValGlyGlyTyrLeuProGlyProAlaGlyValAlaArgGlyPheArgGly 1760  
Db 5503 AGCGTGGTGTCCGAGGCGCCTCTGAAGACAAGGCTCCGTGGCCGCGTAGATTCGAGGC 5562  
Qy 1761 CysLeuGlnGlyValArgValSerAspThrProGluGlyValAsnSerLeuAspProSer 1780  
Db 5563 TGCATGCAAGGAGTAGAGATGGAGGGGGAGCGCCACCAACGTCCGACCCCTGAACATGAC 5622  
Qy 1781 HisGlyGluSerIleAsnValGluGlnGlyCysSerLeuProAspProCysAspSerAsn 1800  
Db 5623 AACGCACTCAAGATCAGGATGAAGACGCGCTGTATGTGAGACAGACCTCTGATCCTGAGC 5682  
Qy 1801 ProCysProAlaAsnSerTyrCysSerAsnAspTrpAspSerTyrSerCysSerCysAsp 1820  
Db 5683 CCTGTGCCCCCAATAGCCGCGTCCACGACGCGCTGGAGAGACTACAGCTGCGTCTGTAC 5742  
Qy 1821 ProGlyTyrTyrGlyAspAsnCysThrAsnValCysAspLeuAsnProCysGluHisGln 1840  
Db 5743 AAAGGGTACCTTGGAAATAACTGTGTGATGCTGTCACTGAACCCCTGGAGAAACATG 5802  
Qy 1841 SerValCysThrArgValProSerSerAlaProHisGlyTyrThrCysGluCysProProAsn 1860  
Db 5803 GGGGCTGCGTGGCGGTCCCGCGCTCCCGGAGGGCTACGTTGGATGTGGGGCCAGT 5862  
Qy 1861 TyrLeuGlyProTyrCysGluIleThrArgIleAspGlnProCysProArgGlyTyrTrpGly 1880  
Db 5863 CACATACGGGGCGGTACTGTGAGAAACAATCGACCTTCCGTGCCCAAGGCTGTGGGGG 5922  
Qy 1881 HisProThrCysGlyProCysAsnCysAspValSerIleGlyPheAspProAspCysAsn 1900  
Db 5923 AAACCCGCTGTGGACCTTGCCACTGTCCGTCAGCAAAAGGCTTGAACCCGACTGTAT 5982  
Qy 1901 LysThrSerGlyGluCysHisCysGlyGluAsnHisTyrAlaGlyProProGlySerProThr 1920  
Db 5983 AAGACCAACGGCCAGGTGCAATGACAAAGAAATTACTACAGCTCTTACGCCAAGACACC 6042  
Qy 1921 CysLeuLeuCysAspCysTyrProThrGlySerLeuSerIleArgValCysAspProGluAsp 1940  
Db 6043 TGTGTGCTCCCTGAGACTGTTCCCTCCCATGCTCCACAGCCGCACTTGGCAATGGCCACC 6102  
Qy 1941 GlyGlnCysProCysLysProGlyValIleGlyArgGlnCysAspArgCysAspAsnPro 1960  
Db 6103 GGGCAGTGTGTGCTGCAAGCCCGGGGTATATCCGGCCGCGCAGTCAACCGTGGACAACCCG 6162  
Qy 1961 PheAlaGluValIleThrAsnGlyCysGluValAsnTyrAspSerCysProArgAlaIle 1980  
Db 6163 TTTGCCGAGGTGCACACGCTCGGTGTGAGATCTACAAATGGCTGTGCCAAGACATTT 6222  
Qy 1981 GluAlaGlyIleTyrTrpProArgThrAspPheGlyLeuProAlaAlaIleProCysPro 2000  
Db 6223 GAGGCGGGCATCTGTGGCCACAGACCAAGTTCCGGCAGCCGGCTGCGCTCCATGCCCT 6282  
Qy 2001 LysGlySerPheGlyThrAlaValArgHisCysAspGluHisArgGlyTyrTrpLeuProPro 2020  
Db 6283 AAAGGATCCGTGGAATAAGCGGTCCGACATCGACGCGGAGGAAGGGGTGGCTGCCCA 6342  
Qy 2021 AsnLeuPheAsnCysThrSerIleThrPheSerGluLeuLysGlyPheAlaGluArgLeu 2040

Db	6343	GAGCTCTTAACTATACCAACCATCTCTTGCTGAGACCTCAGGGGCAAGATGAAACTG	6402
Qy	2041	GLnAryaBngIuSerGlyLeuApsSerGlyYArgSerGlnIuLeuAlaLeuLeuLeuAry	2060
Db	6403	AGCCGCATATGAGACGCGGAGCGAGCGCCAGGCGCTCGACGTGGTGAAGGCGCTGCG	6462
Qy	2061	AsnAlaThrglnnHsThrAlaGlyTyrPhenGlySerAspValIysValAlaTyrGlnLeu	2080
Db	6463	AGTGTACACAGCACACGGGACGCTCTTGGCAATGACGTGCGCAGCGGCTTACAGCTG	6522
Qy	2081	AlaThrArgLeuLeuAlaHisGlySerThrglnAryGlyPheGlyLeuSerAlaThrgln	2100
Db	6523	CTGGGCGCACGCTCTTACAGCACAGAGAGCTGGCAGAGAGGCTTCACTGGCAGCACGAG	6582
Qy	2101	AspValHisPheThrgIuAsnLeuLeuAryValGlySerAlaLeuLeuApsThrAlaAsn	2120
Db	6583	GACGCGCACTTTCACGAGAGAGTCATCACTCGGGCAGCGGCTCTGGCCCGCACAC	6642
Qy	2121	LysArgHsIstProGluLeuIleGlnGlnThrgIuGlyGlyThrAlaIstPheLeuGlnHis	2140
Db	6643	AGGGCGCGGTGGAGCAGATCTCAGCGGAGCGAGGCGGCGGACGCAAGCTCTCGCGC	6702
Qy	2141	TyrGlnAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyrLeuSerProPhe	2160
Db	6703	CTCAGAGGCTACTTTCAGACAGTGCACACGCGCAGCGGCGAGAGTACCTTGGCCCTTC	6762
Qy	2161	ThrIleValThrProAsnIleValIleSerValValArgLeuApsIysGlyAsnPheAla	2180
Db	6763	GTCATTCGTCAACCGGCACACATAGGTTCTTGCTGTGCACATCTTTCAGATTCATCTTACG	6822
Qy	2181	GlyAlaIysLeuProArgTyrGlyAlaLeuAryGlyGlyGlnProProApsLeuGlnThr	2200
Db	6823	GGAGCGAGGGTCCGCGCATTCGACACCATCATGAAAGTTCCCGCAGGAGCTGAGATCC	6882
Qy	2201	ThrValIleLeuProGluSerValPheArg-----GlnThrProProValAla	2216
Db	6883	TCCGTCTCTCTTCCGACCGCACTTCTTACAGCACCTGAGAAAGAAAGAGCCCTGCTG	6942
Qy	2217	ArgProAlaGly-----ProGlyGlnAlaGlnIuPro	2227
Db	6943	AGGCGCGGTGCGGAGAGAACACCCCGCAGACACACGCGCGCGGCGCTGGCAGCAGAGG	7002
Qy	2228	GluGluLeuAlaArgArgGlnAryArgHisProGluLeuSerGlnGlyAlaValAla	2247
Db	7003	GAGGCGCCGATCAGACAGCGGAGGCGGACACCTGATGACGCTGCGCAGTTTCGCGTGC	7062
Qy	2248	SerValIleIleTyrArgThrIleAlaGlyLeuLeuProHisAsnTyrAspProApsIys	2267
Db	7063	CTGTGTATCATTTACCGCACCCCTGGGGCAGCTCTGCGCAGCGCTACAGCCCGACCGT	7122
Qy	2268	ArgSerLeuAryValProIysArgProIleIleAsnThrProValIleSerIleSerVal	2287
Db	7123	CGCAGCTCCGGTGTGCTCACCGGCCCATCATTTAATCCCGAGTGGTGAAGCAGCTGTG	7182
Qy	2288	HisAspApsGluLeuLeuLeuProAryAlaLeuApsIysProValIthValGlnPheArg	2307
Db	7183	TACAGCGAGGGGCTCCGCTCCCGAGACCCCTGGAGAGGCGCCGTCGTGGAAATTCGCG	7242
Qy	2308	LeuLeuGlnThrgIuGlnIuAryThrgIuIysProIleCysValPheTrpAsnHisSerIleLeu	2327
Db	7243	CTGTGTGAGGTGAGGAGCGAACCAACCTGTGTGCGCTGTTCGGAACCACTCTCTGGCC	7302
Qy	2328	ValSerGlyThrgIuGlyTyrTrpSerAlaArgGlyCysGlyValValPheArgAsnGlySer	2347
Db	7303	GTTGTGTGGAGCGGAGGGGTGTCTGCGCGGGGCTGCGAGCTCTCTGTCCAGAAACCGACA	7362
Qy	2348	HisValSerCysGlnCysAsnHisMetThrSerPheAlaValLeuMetAspValSerArg	2367
Db	7363	CATGTCCGCTGCCAGTGCACACACACAGCAGCTTTCGGGTGCTCAGATATCTCAGG	7422
Qy	2368	ArgGluAsnGlyGlyIleLeuProLeuIysThrIleuThrTyrValAlaLeuGlyValThr	2387
Db	7423	CGTAAAGACGGGAGGTCTGCTGTGAAATATGTGCACCTAATGCGCGCTGTGCTCTTGTCA	7482

QY 2388 LeuAlaAlaLeuLeuThrPhePheLeuThrLeuLeuArgIleLeuArgSerAan 2407  
 DB 7483 CTGGGAGCCCTGCTGGGCTTCTGCTCTGAGCTGGCTGGCAGTGCCTCCAAAC 7542  
 QY 2408 GlnHicGlyIleArgArgLeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeu 2427  
 DB 7543 CTGCACAGCATTCACAGACCTTCGCGCTGCTCTCTCTCAAGCTGGGCTTCG 7602  
 QY 2428 LeuGlyIleAsnGlnAlaAspLeuProPheAlaCysThrValIleAlaIleLeuLeuHis 2447  
 DB 7603 ATTGGATCAACCAAGCAAGAAACCGTTCTGTGACAGTGTTCCTCTCCAC 7662  
 QY 2448 PheLeuTyrlLeuCysThrPheSerTrpAlaLeuLeuGlnAlaLeuHisIleuTyrlArgAla 2467  
 DB 7663 TACATCTACATGAGCACTTTCCTGACCTGTGAGAGCCTGCAATGCTACCCCAAG 7722  
 QY 2468 LeuThrGlnValArgAspValAsnThrGlyProMetArgPheTyrlTyrlMetLeuGlyTrp 2487  
 DB 7723 CTACCGAGGTGGCAACATCGACAGGGCCCAATGGGCTTACTACAGCTCGGGCTGG 7782  
 QY 2488 GlyValProAlaPheIleThrGlyLeuAlaValGlyLeuAspProGlnGlyTyrlGlyAan 2507  
 DB 7783 GGATCCCGGCTTGTGACAGAGCTGGCGCTGGCTGAGCCCAAGGCTACGGGAAC 7842  
 QY 2508 ProAspPheCysTrpLeuSerIleTyrlAspThrLeuIleTrpSerPheAlaGlyProVal 2527  
 DB 7843 CCGGACTTCTGCTGGCTGCTGCTTCAAGACACCTGATTTGAGCTTTCGGGGCCCAAC 7902  
 QY 2528 AlaPheAlaValSerMetSerValPheLeuTyrlIleLeuAlaIleArgAlaSerCysAla 2547  
 DB 7903 GGAGCTGTATATATCATCAACAGTCATCTTGCTATCTCAAGGTTCTCTCCAA 7962  
 QY 2548 AlaGlnArgGlnGlyPheGlnGlyGlyGlyProValSerGlyLeuGlnProSerPheAla 2567  
 DB 7963 AGAAGACACATTAATATGGGAAAAAGGATGCTCCCTCGTGAGAACCGATTCCTC 8022  
 QY 2568 ValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeuSerValAsnSerAspThr 2587  
 DB 8023 CTGCTCTGCTCATGAGCCGACCTGCTGCTGGGCTGCTGCTGAGAACCCGAGTGA 8082  
 QY 2588 LeuLeuPheHisTrpLeuPheAlaThrCysAsnCysIleGlnGlyProPheIlePheLeu 2607  
 DB 8083 CTGAGCTTCTACTACTCTTCTGCTATCTTCAAGCGGCTTACAGGCCCCCTTCCTCTT 8142  
 QY 2608 SerTyrlValValLeuSerIleGlyValArgValAlaLeuLys---LeuAlaCysSerArg 2626  
 DB 8143 TTCACCTGCTGCTCAACAGAGGCTCCGAGAGACCTGAAAGGCGGCTCGCGGAGG 8202  
 QY 2627 LysProSerProAspProAlaLeuThrTrpLysSerThr---LeuThrSerSerTyrlAan 2645  
 DB 8203 AAGCTGCACTGAGAGCTCCGACACACAGGAGCCACTGTGAGCGGCTCCCTCAAC 8262  
 QY 2646 CysProSerProTyrlAlaAspGly---ArgLeuTyrlGlnProTyrlArgSerAla 2663  
 DB 8263 TGCACACACACCTTCGGTGCAGGCGCTGACATGCTGCACACACTTGGCGAGTCCACC 8322  
 QY 2664 GlySerLeuHisSerThrSerArgSer--- 2672  
 DB 8323 GCCTCGCTGAGACATCGTCAGGAGTGAAGGATCCAGAAAGCTCGGCGTCTGAG 8382  
 QY 2673 ---GlyLysSerGlnProSerTyrlIleProPheLeuLeuArgGlu 2686  
 DB 8383 CTGGTAGAGGAGCAGCAGCAGAGCAGACGCGTCCCTCAATGCCAGAGCTGCAAGAGT 8442  
 QY 2687 GluSerAlaLeuAsnProGlnGlyGlnGlyProGlnGlyLeuGlyAspProGlySerLeuPhe 2706  
 DB 8443 ---CCCCCTGGC--- 8451  
 QY 2707 LeuGlnGlyGlnAspGlnGlnHisAspProAspThrAspSerAspSerAspLeuSerLeu 2726  
 DB 8452 ---CACGATTCGAGACTCAGATGAGCGAGCTGTCCCTG 8484

QY 2727 GluAspAspGlnSerGlySerTyrlAlaSerThrHisSerSerAspSerGlnGlnGln 2746  
 DB 8485 ---GATGAGCAGAGAGAGCTTACTACCTCCCTCACTGCTGACAGAGAGCATGAG 8541  
 QY 2747 GlnGlnGlnGlnGlnGlnAlaAlaPheProGlnGlnGlnGlnTrpAspSerLeuLeuGly 2766  
 DB 8542 GTGGAGCTGAGAGAAA---TGGACCCGGCCAGGGG 8577  
 QY 2767 ProGlnAlaGlnArgLeuProLeuHisSerThrProLys---AspGlyGlyProGlyPro 2785  
 DB 8578 ---GCGTCCACAGACCCCAAAAGGAGAGCTGTGCAACAC 8619  
 QY 2786 GlyValAlaProTrpProGlnAspPheGlyThrAlaLysGlnSerSerGlyAsnGly 2805  
 DB 8620 GTTCCGGCGGCTGGCCGACAGAGCCTGTGAGAGTGAAGTGAAGAGACCCAGCGGC 8679  
 QY 2806 AlaPro---GlnGlnArgLeu 2811  
 DB 8680 AAGCCCGCTGAGAGTGAAGACCAAGTCAAGCTGAGCTGACCCGAGAGAGCAGGCG 8739  
 QY 2812 ArgGlnAsnGlyAspAlaLeuSerArgGlnGlySerLeuGlyProLeuProGlySerSer 2831  
 DB 8740 ACTCACCGTGAAGAGTACCCCGGACAGAGAGCGGGGCGAGCCAGCTTCTAGC 8799  
 QY 2832 AlaGlnPro---HisLysGlyIleLeuLysLysCysLeu---Pro 2845  
 DB 8800 AACCAGCCCGCAGACAGAGAGCATCTTGAATAATGAATCACTTACCCCGCGCTG 8859  
 QY 2846 ThrIleSerGlnLysSer---SerLeuLeuArgLeuProLeuGlnGlnCysThrGly 2863  
 DB 8860 ACGTGAAGAGACACACCTGAAGGGCCGGCTCCGGAGAACTGGCCGACTGTAGCAG 8919  
 QY 2864 SerSerArgGlySerSerAlaSerGlnGlySerArgGlyGlyPro---Pro 2879  
 DB 8920 AGCCCAACATCTCGCGACGCTTCCCTGGGCTGTGGCGGCCGACATGCATCACA 8979  
 QY 2880 ProArgProProProArgGlnSerLeuGlnGlnGlnLeuAsnGlyValMetProIleAla 2899  
 DB 8980 GTCAAGAGCCCTGGAGGAGGAGCGGGCGTGAACCTTCAACGGGGTG---GCC 9030  
 QY 2900 MetSerIleLysAlaGlyThrValAspGlnLysSerSerGlySerGln 2915  
 DB 9031 ATGATGTGCGCATGTGGAGCGCCAGGCCGATGCTCGACTGTAG 9078

RESULT 13  
 US-10-225-567A-655  
 ; Sequence 655, Application US/10225567A  
 ; Publication No. US20030113798A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lifespan Biosciences  
 ; APPLICANT: Brown, Joseph P.  
 ; APPLICANT: Burner, Glenna C.  
 ; APPLICANT: Roush, Christine L.  
 ; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
 ; FILE REFERENCE: 1970-4-4  
 ; CURRENT APPLICATION NUMBER: US/10/225,567A  
 ; PRIOR FILING DATE: 2001-12-19  
 ; PRIOR FILING DATE: 2000-12-19  
 ; NUMBER OF SEQ ID NOS: 2292  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 655  
 ; LENGTH: 11965  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-225-567A-655

Alignment Scores:  
 Pred. No.: 0 Length: 11965  
 Score: 7836.50 Matches: 1570  
 Percent Similarity: 66.71% Conservative: 424  
 Best Local Similarity: 52.53% Mismatches: 817  
 Query Match: 50.41% Indels: 179

DB: 15 Gaps: 48  
 US-09-916-849a-3 (1-2923) x US-10-225-567A-655 (1-11965)  
 Oy 28 ProLeuLeuGlyAAspGln-----ValGlyProCysArg 38  
 Db CTTCTTGGCGGGGCGAGACGACCTTTCCAAAGAGTGTCTGTACACGAGGGCTGTCTCT 759  
 Oy 39 SerLeuGlySerArgGlyArgGlySerSer----- 48  
 Db 760 CAG-----GGGTCCCGGGCTCGGGGAGACAGCTCCGCCCTCCCTCAGACTTTTGAATTGGCC 816  
 Oy 49 -----GlyAlaCysAlaProMetGlyTyrPLeuCysProSerSerAlaSeran 64  
 Db 817 ACCACGGTCCCAAGCCGGGTCTCTCCACCGGAAACCTGGAGACAGCTCCCAAGAG 876  
 Oy 65 LeuTyrPLeuTyrThrSerArgCysArgAspAlaGlyThrGluLeuThrGlyHisLeuVal 84  
 Db 877 ---TGGGACCCGGCGCTCTGTGGGGAATTATGGCAACG-----GAGCAAGGGTCTC 927  
 Oy 85 -----ProHisAlaAspGlyLeuArgValTyrCysProGluSerGlyAlaHis 100  
 Db 928 AGGGCGAGAGAGCCACGACAT-----CCGGAGCAAGAAAGAGACAG 966  
 Oy 101 IlePro-----LeuProProAlaProGluGlyCys 110  
 Db 967 CCCCCCGGGAACGTCTTCCAGGGGCTCGGAGATCTGGCCCGAGCTG--GATTCAGCA 1025  
 Oy 111 ProTyrSerCysArgLeuLeuGlyTyrIleGlyHisLeuSerProGluGlyTyrLeuThr 130  
 Db 1026 CCAACGACGCGGAGAGACAGCTCTGTGATCAGGTTCA---GCACCCGACAGTCTCGACA 1082  
 Oy 131 LeuProGluGluHisProCysLeuValProArg-----LeuArgCysGlnSerCys 148  
 Db 1083 GCTCCCGAGCGCGCGCCAGAGGAGATGCGCTCCGGGGTCTCTTCGGTGGCGGCTTC--- 1139  
 Oy 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLeuArg 161  
 Db 1140 ---CTCCCGAGCGCGCGCGCGCGCTCCCGGAGCTCCCGCCGCTCTTGAAGCCAGG 1196  
 Oy 162 SerProGluGluSerLeuGlyGlyArgArgGlyValArgAlaValAsnThrAlaProGlnPhe 181  
 Db 1197 AAAGTAACCTCGGCGAACCAGGACCGCTTCTCGCGCGCGCAACCGCACCGCGAGTTT 1256  
 Oy 182 GlnProProSerTyrGlnAlaThrValProGluAsnGlnProAlaGlyThrProValAla 201  
 Db 1257 CCGCAGTACAACTACACAGACGCTGGTCCGAGAAATGAGAGCAGAGGACCGCGGTGCTA 1316  
 Oy 202 SerLeuAlaGlyAlaIleAspProAspGluGlyValAlaGlyValArgLeuGluTyrThrMetAsp 221  
 Db 1317 CGCGTGGTGTCTCAAGACCCCGGACCGCGCGAGCGCGCGCTTGTACTTCGCTGGCG 1376  
 Oy 222 AlaLeuPheAspSerArgSerAsnGlnPheSerLeuAspProValThrGlyAlaVal 241  
 Db 1377 GCACTCATAGAAACCGCTCGCTGAGCTGTTCAGATTCGACCCGACAGCGCGCTTATTC 1436  
 Oy 242 ThrThrAlaGluGluLeuAspArgGluThrLysSerThrHisValPheArgValThrAla 261  
 Db 1437 CGTACCGGCGGAGCTCTGGACCGCGAGCAGCAGGACGTCATCTGCTGTGTGACCGCG 1496  
 Oy 262 GlnAspHisArgLysMetProArgArgSerAlaAlaValThrLeuThrIleLeuValThrAsp 281  
 Db 1497 CAGGACCAACCGGGTCCCGCGCTCTCGGCGACACAGATGTGGCGGTGACATACCCAGC 1556  
 Oy 282 ThrAsnAspHisAspProValPheGluGlnGlnGluTyrLysGluSerLeuArgGluAsn 301  
 Db 1557 CGCAACGACCACTCCCGCGTTTATGAGCAAGCGCAGTACCGGAGAACCTTTGCGAGAAAT 1616  
 Oy 302 LeuGluValGlyTyrGluValLeuThrValArgAlaThrAspGlyAspAlaProProAsn 321  
 Db 1617 GTGAGAGAGGGGCTACCTTATCTGCAAGCTGCTGCACTGACGCGGAGACCGCGCCCAAC 1676  
 Oy 322 AlaAsnIleLeuTyrArgLeuLeu-----GluGlySerGlyGlySerProSerGluVal 339

Db 1677 GCCAACCTCGCGCTACCGCTTGTGGGGCCGCCAGCTGCGCGGCTGCAAGTCCCGCC 1736  
 Oy 340 PheGluIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGluGlu 359  
 Db 1737 TTGAGATTGATTCACGCTCCGGCTCTATCAGACACAGCGGCGCAGTGAACCGGACAC 1796  
 Oy 360 ValGluSerTyrGlnLeuThrValGluAlaSerAspGlnGlyArgAspProGlyProArg 379  
 Db 1797 ATGAAACCTATAGAGCTGTGTGTGGAGACGACGACGAGGCGAGAAACCGGGCGCGGC 1856  
 Oy 380 SerThrThrAlaAlaValPheLeuSerValGluAspAspAsnAspAlaProGlnPhe 399  
 Db 1857 TCGGCACCTGTCGGGTACATATGCTGTCTAGACGAGACGAAACGAAATCTCTCACTTC 1916  
 Oy 400 SerGluValArgTyrValAlaGlnValArgGluAspValThrProGlyAlaProValLeu 419  
 Db 1917 AGCGAAGACCTTACGTGGCGGAGTGGCGGAGAGATGTGCGCCGCCACACAGTCTGTCTG 1976  
 Oy 420 ArgValThrAlaSerAspArgAspLysGlySerAsnAlaValAlaHisTyrSerIleMet 439  
 Db 1977 CGCGTCAAGCGCACAGTACCGGAGACAGGACGCAACGGAATTGTGTGACTTACATCATC 2036  
 Oy 440 SerGlyAsnAlaArgGlyGlnPheTyrLeuAspAlaGlnThrGlyAlaLeuAspValVal 459  
 Db 2037 AGTGCATATACCGTGTGACACTTGTCCATCGACAGCTCAGCTGCGAGATCAGGTGTG 2096  
 Oy 460 SerProLeuAspTyrGluThrThrIleGluGluTyrThrLeuArgValArgAlaGlnAspGly 479  
 Db 2097 GCACCTCTGAGACTTCGAGGAGAGAGAGATGCTTGGCCATCAGGCGCGAGATCT 2156  
 Oy 480 GlyArgProProLeuSerAsnValSerGlyLeuValThrValGlnValLeuAspIleAsn 499  
 Db 2157 GCGCGGCGCACCGCTGTCCAACACACGCGCGCTGGCGAGCATTCAGGTGTGACATCAT 2216  
 Oy 500 AspAsnAlaProIlePheValSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
 Db 2217 GACCAATCTCTATTTTGTGACGACCGCTTCAAGTTCTGTCTTGGAAATGCTCTCC 2276  
 Oy 520 LeuGlyTyrValLeuValLeuHisValGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 Db 2277 TTGGGTACTAGATCATTCACATTCACAGGAGTCAATGACAGCATCGGAGAAATGCCAGA 2336  
 Oy 540 LeuGluTyrArgLeuAlaGlyValGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
 Db 2337 TTGAGATATCTCCAACTGGGTGTGACACTGATCTCTTTGTGATTAACGCGCACT 2396  
 Oy 560 GlyTyrIleSerValAlaAlaGluLeuAspArgGluGluValAspPheTyrSerPheGly 579  
 Db 2397 GCGTGGGTCTCTGTGAGTGTCCCTTGGACCGTGAAGCTGTGAGACATTACTTCTTGTGT 2456  
 Oy 580 ValGluAlaArgAspHisArgLysThrProAlaLeuThrAlaSerAlaSerValSerValThr 599  
 Db 2457 GTGAGGCTGAGACATGCTGCTACCCCACTCTCTGCTCAGCAGGTGACCTGACT 2516  
 Oy 600 ValLeuAspValAsnAspAsnAsnProThrPheThrGlnProGluTyrThrValArgLeu 619  
 Db 2517 GTGCTGAGCGTTATATGCAATCGGCTGAGTTTCAATGAAAGAGTACACCTACGACTG 2576  
 Oy 620 AsnGluAspAlaAlaValGlyThrSerValValThrValSerAlaValAspArgAspAla 639  
 Db 2577 AATAGAGATGAGCTGTGGGACCAAGTGTGTGAGCGGTGACCGGAGTAAACCGGATGCTC 2636  
 Oy 640 HisSerValIleThrTyrGlnIleThrSerGlyAsnThrArgAsnAspPheSerIleThr 659  
 Db 2637 AACGTGCATCAGCTACATACAGTACAGCGGCGCAACCCCGAATTCCTTTGCCATCAGC 2696  
 Oy 660 SerGlnSerGlyGlyLeuValSerLeuAlaLeuProLeuAspTyrLysLeuGluArg 679  
 Db 2697 ACCCAGGGGGGTGTGGGTGTGTGATCTTGTGCTGCACTGAGCTACAGAGCAAGAAACGC 2756  
 Oy 680 GlnTyrValLeuAlaValThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699

Dh 2757 TACTTCAAGCTGTACTAATGATCGACCGGCGCTTCATGATCTGATGTCAC 2816  
Qy ValAsnValThrAspAlaSerThrHisArgProValPheGlnSerSerHisArgThrVal 719  
Db 2817 ATCAACATTCACAGATGCGCACTCATCGCGGCTTTTCAAGATGCCACTACTCATG 2876  
Qy 720 AsnValAsnGluAspArgProAlaGlyThrThrValValLeuIleSerAlaThrArgGlu 739  
Db 2877 AGTGTGAATGAAATCGCGCAATGGGTAGCACATAGTGTCATCACTGCTCTGATGAT 2936  
Qy 740 AspThrGlyGluAsnAlaArgIleThrThrPheMetGluAspSerIleProGlnPheArg 759  
Db 2937 GACGTGGTGAAGATGCTCGATCACTATCTCTGGAGAGAACACCGCCCACTTCGCG 2996  
Qy 760 IleAspAlaAspThrGlyAlaValThrThrGlnIleGluLeuAspArgGluVal 779  
Db 2997 ATTGATCAAGATCTCAGAGCACTTACATTAACAGCCCATTAAGCTTAGAGACCAAGTG 3056  
Qy 780 SerThrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGlnIleSerAspThrThr 799  
Db 3057 ACCTACACCTCGCTATCACAGCTCGGAGACATGGCATCCCAAGAGGACAGACACTACT 3116  
Qy 800 TyrLeuGluIleLeuValAsnAspValAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
Db 3117 TATGTGAGAGGTGATGTCATAGACGTAAATGACATGCTCCACATTTGGGCTCCAC 3176  
Qy 820 TyrGlnIleSerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
Db 3177 TATACAGGCGTGTCTCTGAGAGATGCCCACTTTCACAGTGTCTGCAATTTCTACGCC 3236  
Qy 840 ThrAspArgAspSerGlyLeuAsnGlyArgValPheTyrThrPheGlnIleGlyAspAsp 859  
Db 3237 ACTGACCGGAGTGTCTCATGCGCAATGCGGCGGTCCAGTACACTTCCAGAAATGGAGAT 3296  
Qy 860 GlyAspGlyAspPheIleValGlnSerThrSerGlyIleValArgThrLeuArgGluLeu 879  
Db 3297 GGGGATGAAATTTTACCATGAGCCCACTGTGAATGTCTCTACAGTAAAGCGGCTA 3356  
Qy 880 AspArgGluAsnValAlaGlnTyrValLeuArgAlaTyrAlaValAspGlyMetPro 899  
Db 3357 GACCGGAGAGATACAGTGTAGTGTGATGATGCTGCTACAGAGTGAAGAGGTGTGCC 3416  
Qy 900 ProAlaArgThrProMetGluValThrValThrValLeuAspValAsnAspAspPro 919  
Db 3417 CCACCTCGGAGCTCCAGTCAGTATCCAGGTGATGATGAGTGAAGCAAGCAAGCACT 3476  
Qy 920 ValPheGluGluAspGluPheAspValPheValGluGluAsnSerProIleGlyLeuAla 939  
Db 3477 GTCTTCCAGCTGAGAGTGTGAGGTGCGGCTGAAGAGATACCATTTGGGCTCAGTG 3536  
Qy 940 ValAlaArgValThrAlaThrAspProAspGluGlyThrAsnAlaGlnIleMetTyrGln 959  
Db 3537 GTGGCCAGATCACTGCAAGTGAAGCCCTGACGAAGGCCCAATGCCCATTAATGATACAG 3596  
Qy 960 IleValGluGluAsnIleProGluValPheGluLeuAspIlePheSerGlyGluLeuThr 979  
Db 3597 ATCTGTGAGGGGAAATCTCCCTGAGCTGTTCAAATGACATCTTCTGAGGAATCAGC 3656  
Qy 980 AlaLeuValAspLeuAspArgGluAspArgProGluTyrValLeuValIleGlnAlaThr 999  
Db 3657 GCATCTATTAAGCTTAAGCTTAAGGCTGCGCAAGAAATATGATTTGGTGAGGCGCA 3716  
Qy 1000 SerAlaProLeuValSerArgAlaThrValHisValArgLeuLeuAspArgAsnAspAsn 1019  
Db 3717 TCTCTCTCTTTGGTCAAGCGGCGCACTGTGCACTCGCGCTGGTGAACGAATGACAC 3776  
Qy 1020 ProProValLeuGluAsnPheGluIleLeuPheAsnMetTyrValThrAsnArgSerSer 1039  
Db 3777 AGCCCTGTGCTCAACATCTTCAGATCTTCAACAACATATGATCAACCGTTTCAGAC 3836  
Qy 1040 SerPheProGlyGlyAlaIleGlyArgValProAlaHisAspProAlaIleSerAspSer 1059  
Db 3837 ACCTTCCGTGCGGCAATTAATGGCGCATCCCACTTAATGACCCGATGTCTCCGACAC 3896

Qy 1060 LeuThrTyrSerPheGluArgGlyAsnGluLeuSerLeuValLeuAsnAlaSerThr 1079  
Db 3897 CTCTTCTACTCTTTAGCGGTGGCAATGAGCTGCACTGTGTATTCACAGACCACT 3956  
Qy 1080 GlyGluLeuIleuSerArgAlaLeuAspAsnAspArgProLeuGluAlaIleMetSer 1099  
Db 3957 GGGGACCTCGACTCAGCGGAAAGCTAAGCAATTAACCGCCCACTGTGTGCTCCATGTTG 4016  
Qy 1100 ValLeuValSerAspGlyValHisSerValThrAlaGlnCysAlaLeuArgValThrIle 1119  
Db 4017 GTACTGTCAAGATGGCTGCAAGCGTGAAGCGGCAAGTGTGTGCGCGTGTATC 4076  
Qy 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgGluIleuAspMetProGlu 1139  
Db 4077 ATACCGAGAGAGTGTCTGCGCAACAGCTGACCGTGTGCTTGAAGAACATGTGGCAGAG 4136  
Qy 1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnAlaValAlaAlaThrLeuAlaThr 1159  
Db 4137 CGCTTCTGTCAACCGCTGCGGCGCTTCTCAGAGGCGGTGCTCGGTGTCTGACG 4196  
Qy 1160 ProProAspHisValValAlaPheAsnValGlnArgAspThrAspAlaProGlyGlyHis 1179  
Db 4197 CCGCTGAGAGAGCTTTCACTTCAACATCCAGAACGACACAGACGTA---GGGGGACCC 4253  
Qy 1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyGlyPro----- 1197  
Db 4254 GTGCTCAATGTAGTTTCTCGGCGTACGTCACGTCACGTCGAGGCGGCGGCGCTGACAGG 4313  
Qy 1198 ProPheLeuProSerGluAspLeuGlnIleuGlyLeuTyrLeuAsnArgSerLeuLeuThr 1217  
Db 4314 CCTTGTTCAAGCTCCAGAGAGCTGCAAGAGATTTGATGTCGCGGCGGCGGCTGAGG 4373  
Qy 1218 AlaIleSerAlaGlnArgValLeuProPheAspAspAsnIleCysLeuArgGluProCys 1237  
Db 4374 GTTGGCTCTCTGTGCAAGTATGCTCTTGTGAGAACAGAGTGTGCTGGAGAGCCCTGT 4433  
Qy 1238 GluAsnTyrMetArgCysValSerValLeuArgPheAspSerSerAlaProPheIleAla 1257  
Db 4434 GAAACATACATGAATAGCGTGTCCGTCGCTCCGCTTGTGATCGTCCGCGCTTCTGAGCC 4493  
Qy 1258 SerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysPro 1277  
Db 4494 TCGGCTTCAACGCTGTTCGCAACCCATCCAGCCCATGCTGGCGCTGCGCTGCGCTCCG 4553  
Qy 1278 ProGluPheThrGlyAspTyrCysGluThrGluValAspLeuCysTyrSerArgProCys 1297  
Db 4554 CCGGATTCACGGGAGACTTTTCCAGACCGAGCTGACTCTGTCTTCAACCATGT 4613  
Qy 1298 GlyProIleGlyArgCysArgSerArgGluGlyTyrThrCysLeuCysArgAspGly 1317  
Db 4614 CGAACGGGCGAGCTGCGCGCGCGCGGAGGAGGCTACAGCGTGTGCGCGCGCGCGC 4673  
Qy 1318 TyrThrGlyGluHisArgGluValSerAlaArgSerGlyArgCysThrProGlyValCys 1337  
Db 4674 TTACCGGAGAGGACCTGCAAGCTGACACAGAGCGCGCGCTGTGCGCGCGGTGTG 4733  
Qy 1338 LysAsnGlyTyrCysValAsnLeuLeuValGlyPheLeuCysAspCysProSer 1357  
Db 4734 CGGACGGGGGACCTGACACGACGCGCCCAAGCGGCGCTTGTGCTGCAAGTCCCGGCA 4793  
Qy 1358 GlyAsp---PheGluAspProTyrCysGlnValThrThrArgSerPheProAlaHisSer 1376  
Db 4794 GCGGCGGCTTCAAGGGCGCGCGCTCGAGAGTGTGCGCGCTCTTCCGCGCACTTG 4853  
Qy 1377 PheIleThrPheArgGlyLeuArgGluArgPheHisPheThrLeuAlaLeuSerPheAla 1396  
Db 4854 TTGTGATGTTTGTGCGGCTGTGCGGCAATTCACCTTAAGTGTGCTCTGTGTGCGG 4913  
Qy 1397 ThrLeuGluArgAspGlyLeuLeuLeuTyrAsnGlyArgPheAsnGlyIleuAspPhe 1416  
Db 4914 ACAGTGACAGAGAGCGGCGTGTCTTCAACAGGCGCGCTGAACGAGAGACGACTTC 4973





Db 7125 CAGCGGCGCCCTGGGGCTCCCGACGCGGAGCTGGTAGGCACTGGAGAGTAT 7184  
 QY 2145 AAlaSerAlaLeuAlaGlnMetArgHisThrTyrLeuSerProPheThrIleValThr 2164  
 Db 7185 GCGGCCACACTCCCGAGGAATATGGAATCACTGAATCCATGGGGCTGGTAGC 7244  
 QY 2165 ProAsnIleValIleSerValValArgLeuAspIleValAsnPhe-----AlaGlyAla 2182  
 Db 7245 CCAAAATATCATGCTCAGCATTCGACGATGAGACACCCAGTTCCTCCCGGGGGCCCGT 7204  
 QY 2183 LysLeuProArgTyrGluAla-----LeuArgGlyGluGlnProProAsnLeuGluThr 2200  
 Db 7305 GCGTACCTGGCTTACCATTAACCACTTTCGAGGCCAGATCCCTGGATCTCAACACC 7364  
 QY 2201 ThrValIleLeuProGluSerValPheArgGluThrProProValValArgProAlaGly 2220  
 Db 7365 CATGTGTGTGGCTTCCAGTCCCGACGCGCATCCCATCTGAAGTTCTGGCCACAGC 7424  
 QY 2221 ProGlyGluAlaGlnGluProGluGluLeuAlaArgArgGlnArgHisProGluLeu 2240  
 Db 7425 AGCAGCATAGAAAATCCACACCTCAAGTGTGGTCCCGCCACAGCCCGCCAGAGCCA 7484  
 QY 2241 SerGlnGlyGluAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeuLeuPro 2260  
 Db 7485 GAGCCTGGAGTCTCATTAATCATTCCTCGTTTACCGACCTTAGGGGAGCTCTCCCT 7544  
 QY 2261 HisAsnTyrAspProAspIleArgSerLeuArgValProIleValIleAsnThr 2280  
 Db 7545 GCCAGTTCAGAGCAAGACCGAGTGCAGGCTTCTCAAGAACCCCGTCAAGAACTCC 7604  
 QY 2281 ProValValSerIleSerValHisAspAspGluGluLeuProArgAlaLeuAspIle 2300  
 Db 7605 CCGGTGTACGCGTGTGTGTCCACGACGCAACTCTTAAGGGGAGATCTCGAGTCC 7664  
 QY 2301 ProValThrValGlnPheArgLeuLeuGluThrGluGluArgThrIleAspValIle 2320  
 Db 7665 CCATACGCTTAAGTTCGCTGCTGCTACAGACCGAATGAGAGGCGAATCTGTGTG 7724  
 QY 2321 PheTyrAsnHisSerIleLeuValSerGlyThrGlyGlyTyrPheSerAlaArgGly 2340  
 Db 7725 CAGTGGAGCCCACTGGCTGGCGGAGCAGACATGTTGTGTGACAGACGAGACTGGAG 7784  
 QY 2341 ValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisSerThrSerPheAla 2360  
 Db 7785 CTGGTGTACAGGATGGGTCCACGACGCGTGTGTGTGCGACGCGACGAGCACTTGGG 7844  
 QY 2361 ValLeuMetAspValSerArgArgGly-----AsnGlyGluIleLeuProLeuGlyThr 2378  
 Db 7845 GTCTCATGATGCTCTCCCGTGAAGGCTGAGGGGCAACTGTGAGCTGGCTGTG 7904  
 QY 2379 LeuThrTyrValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThrPhePheLeu 2398  
 Db 7905 TTCACCCAGCTGGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 7964  
 QY 2399 ThrLeuLeuArgIleLeuArgSerArgGlnHisGlyIleArgArgAsnLeuThrAlaAla 2418  
 Db 7965 CTGAGCTGGCGAGCTCAAGTCAATGTGCGTGGATTCATGCGAATGGGACGCGCC 8024  
 QY 2419 LeuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAsnGlnAlaAspLeuProPheAla 2438  
 Db 8025 CTGGGGGTGGCAGACTCTCTCTCTCTGTGGGATTCACAGAGCCCAATCAGCTGGTG 8084  
 QY 2439 CysThrValIleAlaIleLeuLeuHisPheLeuTyrLeuCysThrPheSerTrpAlaLeu 2458  
 Db 8085 TGCATCTGAGTGGCATCTCTCTGCACTACTTCTCTCAGACACTTCGGGTGGCTTTC 8144  
 QY 2459 LeuGluAlaLeuHisLeuTyrArgAlaLeuThrGlyValAlaArgAspValAsnThrGlyPro 2478  
 Db 8145 GTGAGAGGGGTGCACTCTACCGCATGCAAGTTGAGCAGCAGCGGAGCCGCGGCC 8204  
 QY 2479 MetArgPheTyrTyrIleMetLeuGlyTyrGlyValProAlaPheIleThrGlyLeuAlaVal 2498  
 Db 8205 ATGGCTTTCATCACTGCGCTGGGCTGGGGCTCTCTGTGTGTGTGTGTGTGTGTGTG 8264

QY 2499 GlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTrpLeuSerIleTyrAspThr 2518  
 Db 8265 GCGCTGGACCTCGAGGGCTATGGAACTTAATCTTGTGTGATCTCACTCAAGAGCC 8324  
 QY 2519 LeuIleTyrPhePheAlaGlyProValAlaPheAlaValSerMetSerValPheLeuTyr 2538  
 Db 8325 CTCATCTGGAGCTTGTGGCCCTGTGTGCTGTGTATGATGAACGGACATGTTT 8384  
 QY 2539 IleLeuAlaAlaArgAlaSerCysAla-----AlaGlnArgGlnGlyPheGluValValGly 2557  
 Db 8385 CTCTCGCTGCGCCGACATCTCTGTCTCACAGGGCGAGGGAGGCC-----AAGAGACC 8438  
 QY 2558 ProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuSerAlaThrTrpLeu 2577  
 Db 8439 TGTGACATGACCTTGCGAGCTCTCTTCGTCGCTTGTGTGTGTGTGTGTGTGTGTG 8498  
 QY 2578 LeuAlaLeuLeuSerValHisSerAspThrLeuLeuPheHisTyrLeuPheAlaThrCys 2597  
 Db 8499 TTTGGGCTCTGCGAGTCAACACAGCATCTTACGCTTCACTACCTCATCTGAGTTC 8558  
 QY 2598 AsnCysIleGlnGlyProPheIlePheLeuSerTyrValValLeuSerIleGluValArg 2617  
 Db 8559 TCGGCTCTCAGGGCTGGCGGTGCTGTCTTCTGTGTCTTAATGACAGATGCTCGG 8618  
 QY 2618 LysAlaLeuLysLeuAlaCys-----SerArgLysProSerProAspProAlaLeuThrThr 2636  
 Db 8619 GCGTCTGTGATCCAGCGCTGTGTGGGAGAGAGGAGCGCTGAGAGGCAAGCGCAGCA 8678  
 QY 2637 LysSerThrLeuThrSerSerTyrHisCysProSerProTyrAlaAspGlyArgLeuTyr 2656  
 Db 8679 CTTGGCTGGGACCTGGGGCTCTTACAAACACGCGTCTCTTGAAGAGTGGGCTCATC 8738  
 QY 2657 Gln---ProTyrGlyAspSerAlaGlySerLeuHisSerThrSerArgSerGlyLysSer 2675  
 Db 8739 CCGATCATCTGGGGGCTTCCACCGTCTCTTGTGAGCAGTGGCCGCTCCGGAGACC 8798  
 QY 2676 GlnPro-----SerTyrIle-----ProPheLeuLeuArgGlu 2686  
 Db 8799 CAGACACAGACAGCAGCGGGCGGAGCTACCTCAGAGGACAAATGCTGTGTCAGCAT 8858  
 QY 2687 GluSerAla-----LeuAsnProGlyGlnGlyProProGlyLeu 2699  
 Db 8859 GCGTCAGCGCTGACCACTGACCAAGCTTCCAGGCTCATGTGGCCCACTGACCTG 8918  
 QY 2700 Gly-----AspProGlySerLeuPheLeuGluGlyGlnAspGlnGln 2713  
 Db 8919 GAGTGGCCATGTCATGAGATGCTGGCGCA----- 8951  
 QY 2714 HisAspProAspThrAspSerAspSerAspLeuSerLeuGluAspAspGlnSerIleSer 2733  
 Db 8952 -----GACTCCGACTGTAAGTACCTGTCTTGTGAGAGAGAGAGTCTCTCC 9002  
 QY 2734 TyrAlaSerThrHisSerSerAspSerGluGluGluGluGluGluGluAla 2753  
 Db 9003 ATTCATCTTGAAGAAAGCAGGACATGCGCGGAGCGGCGCTTCCAGCGCCATCTC 9062  
 QY 2754 AlaPheProGlyGluGlnGlyTyrAspSerLeuLeuGlyProGlyValAlaGluArgLeuPro 2773  
 Db 9063 TGGCCAGAGCCAG-----AGTGAAGGCTCTTC 9092  
 QY 2774 LeuHisSerThrProLysAspGlyGlyProGlyLysValAspProTrpProGly--- 2792  
 Db 9093 ACCCAC-----CCCAAGATGTGATGAGCATGACTCTCTCTCACTGCGCAGCCCTG 9146  
 QY 2793 -----AspPheGlyThr----- 2796  
 Db 9147 GGGAGTGGAGGACAGCCCTGTGTCTGTGCAAGTCTGGGGCTCTGAAGAGCGCTGGGG 9206  
 QY 2797 -----ThrAlaGlyGluSerSerGlyAsnGlyAlaProGluGluLysArgLeuArgGluAsn 2814  
 Db 9207 CTGGAACACGAGAGATGACACTTAACAACACAGCAGACCGGGCCTG---ACCAAT 9263



```

QY 2815 G1AAspAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySerSerAlaGlnPro 2834
Db 9264 GGGGAT-----GAGACTTCTCTG-----GGCCGGCCCGACGCCCGAG 9299
QY 2835 H1sLyG1Y1LeuYsLySvCyS---LeuProThrIleSerGlnYsSerSerLeu 2853
Db 9300 AGGAAAGGCACTCTGAAAGAACCGTTGCATACCCCTGCTGCACAGACCCGAGGTGCC 9359
QY 2854 LeuArgLeuProLeuGlnGlyGlnGlySerSerArgGlySerSerAla----- 2870
Db 9360 CCTGAG-----CTGCTCTGTCCTCGTCAGCCACCTTGGCCACCGTGACGACGCT 9413
QY 2871 -----SerGlnGlySerArgGlyGlyProProProArgProProPro 2884
Db 9414 GCCTTACGCTGCATCTATGCTGGCGGGCGACGCGGACGCTTTCAGACGACGACG 9473
QY 2885 ArgGlnSerLeuGlnGlnGlnLeuAen 2893
Db 9474 CGCTACTCTTCTAGAGAACGCTGAC 9500

RESULT 14
US-10-149-819-42
/ Sequence 42, Application US/10149819
/ Publication No. US20030044913A1
/ GENERAL INFORMATION:
/ APPLICANT: INCYTE GENOMICS, INC.
/ APPLICANT: YUE, Henry
/ APPLICANT: AZIMZAI, Yalda
/ APPLICANT: TANG, Y. Tom
/ APPLICANT: PATTERSON, Chandra
/ APPLICANT: BAUGHN, Mariah R.
/ APPLICANT: LU, Dying Aina M.
/ APPLICANT: SHAH, Puri
/ APPLICANT: LAU, Preeti
/ APPLICANT: AU-YOUNG, Janice
/ APPLICANT: BURROD, Neil
/ TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
/ FILE REFERENCE: PR-0760 PCT
/ CURRENT APPLICATION NUMBER: US/10/149,819
/ PRIOR FILING DATE: 2002-06-10
/ PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
/ PRIOR FILING DATE: 1999-12-10; 1999-12-16
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: PERL Program
/ SEQ ID NO 42
/ LENGTH: 11648
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030044913A1 2847752CBI
US-10-149-819-42

Alignment Scores:
Pred. No.: 0 Length: 11648
Score: 7766.00 Matches: 1561
Percent Similarity: 66.22% Conservative: 421
Best Local Similarity: 52.16% Mismatches: 812
Query Match: 49.96% Indels: 200
Gaps: 49

US-09-916-849A-3 (1-2923) x US-10-149-819-42 (1-11648)
QY 28 ProteuLeuGlyAspGln-----ValGlyProCysArg 38
Db 416 CCTCTTGCCTGGCGACAGGACCTTTCAGAAAGATGCTGTACACAGGGGCTGTCTCT 475
QY 39 SerLeuGlySerArgGlyArgGlySerSer----- 48
Db 476 CAG-----GGGTCCCGGGCTCTGGGAGACGTCGCCCTCTCCCTTCAGACTTTTGATTCGAC 532
QY 49 -----G1yAlaCysAlaProMetGlyTTPLeuCyProSerSerAlaSerAen 64

```

```

Db 533 ACCAGGTCCTCCAAAGCCGGTCTCTCCAGCGGACGCTGGAGACAGGCTCCGCAAAAG 592
QY 65 LeuTTPLeuYTYrThSerArgCysArgAspAlaGlyThrGlnLeuThrGlyHisLeuVal 84
Db 593 ---TGGGACCGCGGGCTCTGTGGGAAATTATGGGCAACG-----GAGCAAGGCT 643
QY 85 -----ProHisIleAspGlyLeuArgValTTPCyProGlnSerGlnAlaHis 100
Db 644 AGGGGAGAGAGACAGACAT-----CCGAGGAGAAAGAGACAG 682
QY 101 ILePro-----LeuProProAlaProGlnGlyCys 110
Db 683 CCCCCGGGGAATGTTCTCCAGGGGCTCGGAGATCTGGCCCGGCTG--GATTACAGA 741
QY 111 ProTTPSerCysArgLeuLeuGlyY1LeGlyGlyHisLeuSerProGlnGlyYsLeuThr 130
Db 742 CCAAGCAGCGGAGAGAGACAGCTCTGCATCAGGTTA---GCACCCCGAGCTCTGGACA 798
QY 131 LeuProGlnGlnHisProCysLeuLeuYsAlaProArg-----LeuArgCysGlnSerCys 148
Db 799 GCTCCCGAGCGGCGCCCAAGGCGATGCGCTCCGGGGTCTTTCGCTGCGCTTC--- 855
QY 149 LysLeuAlaGlnAlaProGly-----LeuArgAlaGlyLeuArg 161
Db 856 ---CTCCGACAGCGCCCGCGCGCGCTCCCGGACCTCCGCGCGCTTCAGAACCCAG 912
QY 162 SerProGlnGlnSerLeuGlyGlyYArgArgYsArgAenValAsnThrAlaProGlnPhe 181
Db 913 AAAGTAACTCGCGCAACCGGCGACGCTTGTGTCGCGCGCAACCGGACCCGACGTTT 972
QY 182 GlnProProSerTYrGlnAlaThrValProGlnAenGlnProAlaGlyTYrProValAla 201
Db 973 CGGACGTAACTAACACCAACCTGCTGCGGAGATGAGGACGACGACGCGGCTGCTA 1032
QY 202 SerLeuArgAlaIleAspProAspGlnGlyGlnAlaGlyYArgLeuGlnTYrThreAsp 221
Db 1033 CGCGTGTCTCAAGACCCGAGCGCGGAGCGCGGCGCTGACTGCTGCTGCG 1092
QY 222 AlaLeuPheAspSerArgSerArgSerArgGlnPhePheSerLeuAspProValThrGlyAlaVal 241
Db 1093 GCACTCATGAAACAGCGCTGCTGCTGAGCTGTTCAGCATGACCCGCAAGAGGCTTATC 1152
QY 242 ThrThrAlaGlnGlnLeuAspArgGlnThrYsSerThrHisValPheArgValThrAla 261
Db 1153 CGTACGGGCGGAGCTCTGAGCCGAGAGCATGAGCCTCACTCTGCTGTGACCCG 1212
QY 262 GlnAspHisGlyMetProArgArgSerAlaLeuAlaThrLeuThrIleLeuValThrAsp 281
Db 1213 CAGACACCGGCTGCGCGCTCTCGGCACACAGATGTGGCCGTACAGTACGCCAG 1272
QY 282 ThrAsnAspHisAspProValPheGlnGlnGlnGlnTYrYsGlnSerLeuArgGlnAen 301
Db 1273 CGCAACGACCACTGCGCGGCTTTTGAAGCAAGCGATACCGGAGACCTTCGAGAGAT 1332
QY 302 LeuGlnValGlyTYrGlnValLeuThrValAlaArgAlaThrAspGlyAspAlaProProAen 321
Db 1333 GTGAGAGAGGCTACCTCATCTGACGTGCTGCACTGACGCGGAGACCCGCCCAAC 1392
QY 322 AlaAsnIleLeuTYrArgLeuLeu-----GlnGlySerArgYsGlySerProSerGlnVal 339
Db 1393 GCCAACCTGCGCTACCGCTTGTGGGGCGCCGACGTCGCGCGCTGCGCGCGCC 1452
QY 340 PheGlnIleAspProArgSerGlyValIleArgThrArgGlyProValAspArgGlnGln 359
Db 1453 TTCCAGATTGATTCACGCTCGCGCTCATGACACAGCGCCGAGTGAGACCGGAGGAC 1512
QY 360 ValGlnSerTYrGlnLeuThrValGlnAlaSerArgGlnGlyYsAspProGlyProArg 379
Db 1513 ATGAAAGACTYAGAGCTGTGTGTGAAGCCAGCAACGAGGACCAAGCCGCGCGCGC 1572
QY 380 SerThrThrAlaAlaValPheLeuSerValGlnAspAsnAspAsnAlaProGlnPhe 399
Db 1573 TCGGCACTGTCGCGCTACATTAATCTGTCTACAGAAACGACATGCTCTCAAGTT 1632

```

QY 400 SerGluValArgTyrValValGlnValArgGluAspValThrProGluValaProValLeu 419  
 Db 1633 AGCGAAGAGCGCTACGGCGCAGAGTCCGAGAGATGCGCCCAACAGTCGTGCTG 1692  
 QY 420 ArgValThrIleSerAspArgAspArgGlySerAsnAlaValAlaHisTyrSerIleMet 439  
 Db 1693 CGGTCACCGGCACTGACCGGAGCAAGCAAGCATGGTGGCTACATCAATCATC 1752  
 QY 440 SerGluValaArgGlyGlnPheTyrLeuAspAlaGlnThrGluAlaLeuAspVala 459  
 Db 1753 AGTGGCAATAGCCGTGACACTTGGCAATGCAAGCTCCTGCGAGATCCAGTGTG 1812  
 QY 460 SerProLeuAspTyrGluThrThrTyrGluTyrThrLeuArgValaArgAlaGlnAspGly 479  
 Db 1813 GCACCTTGCACTTCGAGCAGAGAGAGATATGCTTGGCATCAGGGCGCAGATGCT 1872  
 QY 480 GluArgProProLeuSerAsnValaSerGlyLeuValThrValGlnValLeuAspAlaAsn 499  
 Db 1873 GCGCGGCAACCGCTGTCACACACACGCGCTGCGCATCCAGGTGGTGAATCATCAT 1932  
 QY 500 AspAsnAlaProIlePheValaSerThrProPheGlnAlaThrValLeuGluSerValPro 519  
 Db 1933 GACCAATTCCTATTTTGTGACGACGCTTCCAAATTTCTGTGCTTGGAAATGCTCCC 1992  
 QY 520 LeuGlyTyrLeuValaLeuHisValaGlnAlaIleAspAlaAspAlaGlyAspAsnAlaArg 539  
 Db 1993 TTGGGTCACTCAGCATCCATCCATTCAGGCGATGATGACAGCATGGGGGAATGCGAGA 2052  
 QY 540 LeuGluTyrArgLeuAlaGlyValaGlyHisAspPheProPheThrIleAsnAsnGlyThr 559  
 Db 2053 TTGGAGTACCTCCCTAACTGATGATGCGACCTGATATCTCTTTGTGAAACAGCGCACT 2112  
 QY 560 GlyTyrIleSerValaAlaGluLeuAspArgGluGluValaAspPheTyrSerPheGly 579  
 Db 2113 GCGTGGGTCTCTGTAGAGTGTCCCTGGACCGTGAAGCTGTGAGCACTTCTTTGGT 2172  
 QY 580 ValGluAlaArgAspHisGlyThrProAlaLeuThrAlaSerAlaSerValThr 599  
 Db 2173 GTGAGGCTCGAAGCAATGCTGCTACCCCACTCTGCTCAGCGATGTCACCGTGACT 2232  
 QY 600 ValLeuAspValaAsnAspAsnProThrPheThrGlnProGluTyrThrValaArgLeu 619  
 Db 2233 GTGTGACGCTTAATGACATCGGCTGAGTTCACATGAAGATGACCACTACAGCTG 2292  
 QY 620 AsnGluAspAlaAlaValaGlyThrSerValaThrValaSerAlaValaAspArgAspAla 639  
 Db 2293 AATGAGATGCACTGTGGGCAACAGTGTGTCAGGTCGACCGAGTAAAGCCGTGATGCC 2352  
 QY 640 HisSerValaIleThrTyrGlnIleThrSerGlyAsnThrArgAsnArgPheSerIleThr 659  
 Db 2353 AACAGTCCCACTACGATCAACGATCAACGCGGCAACACCGGAATCCCTTTGCACTCAGC 2412  
 QY 660 SerGlnSerGlyGlyLeuValaSerLeuAlaLeuProLeuAspTyrIleGluGluArg 679  
 Db 2413 ACCGAGGGGGGTGTGGGTGTGGTGTGACTGTGCTGCTGCACTGACATCAAGAGAGAGC 2472  
 QY 680 GlnTyrValaLeuAlaValaThrAlaSerAspGlyThrArgGlnAspThrAlaGlnIleVal 699  
 Db 2473 TACTTCAAGCTGGATCAATCACTGATGACCGAGCCCTTCAATGATCACTGATGAC 2532  
 QY 700 ValAsnValaThrAspAlaAsnThrHisArgProValaPheGlnSerSerHisGlyThrVal 719  
 Db 2533 ATCAATCACTCAGATGCAACATCATCGCGCTCTTTCAAAGTGGCCCACTACTCAGTG 2592  
 QY 720 AsnValaAsnGluAspArgProAlaGlyThrThrValaValaLeuIleSerAlaThrAspGlu 739  
 Db 2593 AGCTGTAAATGAATGCGCCCAATGGGTAGCAACATAGTGTCACTCACTGCTGTATGAT 2652  
 QY 740 AspThrGlyGluAsnAlaArgGlyThrTyrPheMetGluAspSerIleProGlnPheArg 759  
 Db 2653 GACGTGGGTGAATGCTCGATCACTATCTCTGAGAGACAACTGCGCCCAAGTTCCGC 2712

QY 760 IleAspAlaAspThrGlyAlaValThrThrGlnAlaGluLeuAspTyrGluAspGlnVal 779  
 Db 2713 ATTTGATGCAAGACTGAGAGCCATTACAGGCCCACTTATGACTATAGACCAAGGTTG 2772  
 QY 780 SerTyrThrLeuAlaIleThrAlaArgAspAsnGlyIleProGluLysSerAspThrThr 799  
 Db 2773 ACCATCACTTGGCTATACACTGCGACAAATGGATGCCACAGAGGCAAGCACTACT 2832  
 QY 800 TyrLeuGluIleLeuValaAsnAspValaAsnAspAsnAlaProGlnPheLeuArgAspSer 819  
 Db 2833 TATGTGAGGTATAGTGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 2892  
 QY 820 TyrGlnGlySerValTyrGluAspValProProPheThrSerValLeuGlnIleSerAla 839  
 Db 2893 TATACAGGGCTGTGCTGAGAGATGCCCACTTTTACACAGTGTCTGTGAGATTCACGCC 2952  
 QY 840 ThrAspArgAspSerGlyLeuAsnGlyArgValaPheTyrThrPheGlnGlyValaAspAsp 859  
 Db 2953 ACTGACCGGATGCTCAGTCCATGCGCAGATGCGGCTGCGATCATTTCAGATGTGAAGT 3012  
 QY 860 GluAspGlyAspPheIleValaGluSerThrSerGlyIleValaArgThrLeuArgArgLeu 879  
 Db 3013 GGGGATGAGATTTTACATGATGAGCCCACTTGGAATTTGTCGATACATGAGCGGCTA 3072  
 QY 880 AspArgGluAsnValaAlaGlnTyrValaLeuArgAlaTyrAlaValaAspArgMetPro 899  
 Db 3073 GACCGGAGGCGATACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3132  
 QY 900 ProAlaArgThrProMetGluValaThrValaThrValaLeuAspValaAsnAspProPro 919  
 Db 3133 CAATCCGAGCTCCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3192  
 QY 920 ValPheGluGlnAspGluPheAspValaPheValaGluGluAsnSerProIleGlyLeuAla 939  
 Db 3193 GTCTTCCAGCTGAGAGATTTGATGAGTGGGGAAGAAATAGCATTTGTGGCTCAGTG 3252  
 QY 940 ValAlaArgValaThrAlaThrAspProAspGlyGlyThrAsnAlaGlnIleMetTyrGln 959  
 Db 3253 GTGGCCCAATACATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3312  
 QY 960 IleValaGluGlyAsnIleProGluValaPheGlnLeuAspIlePheSerGlyGluLeuThr 979  
 Db 3313 ATCTGAGAGGGAATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3372  
 QY 980 AlaLeuValaAspLeuAspTyrGluAspArgProGluTyrValaLeuValaIleGlnAlaThr 999  
 Db 3373 GCACTCATTTGACCTTATGATGAGCTGCGCAAGATATGATGATGATGATGATGATGATGAT 3432  
 QY 1000 SerAlaProLeuValaSerArgAlaThrValaHisValaArgLeuAspArgAsnAspAsn 1019  
 Db 3433 TCTGCTCTTTGTGTGACCGGCGCACTGTGACGTCGCTGTGTGACCAAGATACAAAC 3492  
 QY 1020 ProProValaLeuGlyAsnPheGluIleLeuPheAsnAsnTyrValaThrAsnArgSerSer 1039  
 Db 3493 AGCCCTGTCTCAACAATCTTCAGATCTCTTTCACAACTATGATTCACACCGTTCAAGC 3552  
 QY 1040 SerPheProGlyGlyAlaIleGlyArgValaProAlaHisAspProAspIleSerAspSer 1059  
 Db 3553 ACCTTCCCGTGGGATATATGGCGCATGCCAGCTTATGACCCGAGATGTCCTCCACAC 3612  
 QY 1060 LeuThrTyrSerPheGluArgGlyValaGluSerLeuValaLeuLeuAsnAlaSerThr 1079  
 Db 3613 CTCTTCTATCTCTTTGAGGTGTGACATGATGATGATGATGATGATGATGATGATGATGAT 3672  
 QY 1080 GlyGluLeuValaLeuSerArgAlaLeuAspAsnAsnArgProLeuGluValaIleMetSer 1099  
 Db 3673 GGGAGTGTGCACTGACCGGAAGCTTAAAGATTAACCGCCATGCTGTGCTCCATGTTG 3732  
 QY 1100 ValLeuValaSerAspGlyValaHisSerValaThrAlaGlnCysAlaLeuArgValaThrIle 1119  
 Db 3733 GTGAGCTGCACAGATGGCTGTGACAGCGGCGAGTGTGTCTGTGCTGTGTGATC 3792  
 QY 1120 IleThrAspGluMetLeuThrHisSerIleThrLeuArgLeuGluAspMetSerProGlu 1139

```

Db      3793 ATACGAGAGAGTTGCGGCCAACAGCTGACCGTGGCCCTTGAGAACATGTGGCAGAG 3852
Qy      1140 ArgPheLeuSerProLeuLeuGlyLeuPheIleGlnIleValAlaIleThrLeuAlaThr 1159
Db      3853 CGCTTCGTGACACCGCTGCGGCGCTTCTCGAGGGCGGTGGGTGGCTCGGCTGACG 3912
Qy      1160 ProProAerPheIleValValPheAsnValGlnArgPheThrAraProGlyGlnIle 1179
Db      3913 CCGCGTAGAGGACGCTTCACTTCAACATCCAGAACAGACAGACGGA---GGGGGACCC 3969
Qy      1180 IleLeuAsnValSerLeuSerValGlyGlnProProGlyProGlyPro----- 1197
Db      3970 GTGCTCAATGTAGATTCTCGGGCTAGCTCAAGTGGGGCCGGGGCGCGCTGGACGG 4029
Qy      1198 ProPheLeuProSerGluAerPheLeuGlnGluArgLeuTyThrLeuAsnArgSerLeuThr 1217
Db      4030 CCTCGGTCAAGCTCCAGAGAGCTGCAGAGACAGTTGATCGTGGCCGGGGCGCGCTGG 4089
Qy      1218 AlaIleSerAlaGlnArgValLeuProPheAerAerAerIleCysValArgGluProCys 1237
Db      4090 GCTCGCTCCCTGCTCGACAGTACCTGCTGACAGACAGTGGCTGGAGAGAGCTGT 4149
Qy      1238 GluAerTyThrArgCysValSerValLeuArgPheAerSerAlaProPheIleAla 1257
Db      4150 GAAAGTACATGAAATGCGGTGCTCGCTCGCTGACCTGCTCGCGCTTCTGCGCC 4209
Qy      1258 SerSerSerValLeuPheArgProIleHisProValGlyGlyLeuArgCysArgCysPro 1277
Db      4210 TCGGCTTCACGCTGTTCCAGCCATCCAGCCACGCTGCGCTGCGCGCTGCGCTCGCG 4269
Qy      1278 ProGlyPheThrGlyAerTyCysGluThrGlnValAerLeuCysValSerArgProCys 1297
Db      4270 CCGGAGATTCACGGGAGACTTTTGCAGACCGAGCTGACCTGCTGCTGCTGCTGCAAC 4329
Qy      1298 GlyProHisGlyArgCysArgSerArgGlnGlyGlyTyThrCysValCysValArgArgGly 1317
Db      4330 CGCAGCGGAGACCGCTGCGCGCGCGCGAGAGAGCTACAGCTGCTGCGCGCGCGCG 4389
Qy      1318 TyThrGlyGlyIleHisCysGlnValSerAlaArgSerGlyArgCysThrProGlyValCys 1337
Db      4390 TTCAACGGAGAGAGACTGCGAGCTGACACCGAGAGCGCGCGCGCTGCGCGCGCTGCG 4449
Qy      1338 LysAerGlyGlyTyThrCysValAerLeuLeuValGlyGlyPheLysCysAerCysProSer 1357
Db      4450 CGCAGCGGGGCGACCTGACCGACCGCGCCAGCGGGCTTTGCTGCGACAGTGGCGGGA 4509
Qy      1358 GlyAer---PheGlyLysProTyTyGlnValThrThrArgSerPheProAlaHisSer 1376
Db      4510 GGGGGGCGCTTCAGAGGCGCGCGCTGCGAGGTGGCTGCGCGCTTCCCGCGCGAGTTG 4569
Qy      1377 PheIleThrPheArgGlyLeuArgGlnArgPheHisPheThrLeuAlaLeuSerPheAla 1396
Db      4570 TTGCTGATGTTTGGCGGCTGCGGCGAGCATTCACCTTACGCTGTCCTCTGCTGCGG 4629
Qy      1397 ThrLysGluArgAerGlyLeuLeuLeuTyArgGlyArgPheAsnGlyLysAerPhe 1416
Db      4630 ACGATGACAGAGCGGGCTGCTTCTACACGGGCGCTGACAGAGAGACAGACCTTC 4689
Qy      1417 ValAlaLeuGluValIleGlnGlnValGlnLeuThrPheSerAlaGlyGlySerThr 1436
Db      4690 CTGGCGCTGGAATCTGAGCTGGCGCAAGTGGCGCTACATATTCACCGGTGATATCCAC 4749
Qy      1437 ThrThrValSerProPheValProGlyGlyValSerAerProGlyIleThrIleAlaGln 1456
Db      4750 ACCGTGTGACGCCACAGTTCAGAGGGGCTTGAAGTGAAGCGCAATGCAATGAGTGCAT 4809
Qy      1457 LeuLysTyTyThrAsnLysProLeuLeuGlyGlnThrGlyLeuProGlnIleProSerGlu 1476
Db      4810 CTGAGATACATACAAACAGCCCGGACAGATGCGCTGAGGGGTGACAGAGGCCCTTCAG 4869
Qy      1477 GlnLysValAlaValValThrValAerGlyCysAerThrGlyValAlaLeuArgPheGly 1496

```

```

Db      4870 GACAAAGTGGCTGTCTAAGCGTGGATGATTGTGATGTGGCGCGGTGCTGCAATTGGT 4929
Qy      1497 SerValLeuGlyAerTySerCysAlaIleGlnGlyThrGlnGlyGlySerTyLysSer 1516
Db      4930 GCTGAGATTGGACATACATCAGCGCGCTGTGTGTGCAACAGACTCCAGAGAGTCC 4989
Qy      1517 LeuAerLeuThrGlyProLeuLeuGlyGlyValProAerLeuProGlyLysPhePro 1536
Db      4990 CTGAGACTGACGGGCGCTTCTTCTGGAGAGGTGTCCCAACTCCCGAGAACATTCGCC 5049
Qy      1537 ValArgMetArgGlnPheValGlyCysMetArgAerLeuGlnValAerSerArgHisIle 1556
Db      5050 GTATCCCAATAGAGACTTACCGCTGTATGCGGAGACTGCACATGATGAGCGCGAGTGG 5109
Qy      1557 AarMetAlaAerPheIleAlaAerAsnGlyThrValProGlyCysProAlaLysPheAsn 1576
Db      5110 GACATGGCGGCTTTGTGTGCATATATGACACATGGCAGCTGCGCCAGCCAGCTTACAC 5169
Qy      1577 ValCysAerSerAerThrCysHisAerGlyTyThrCysValAerGlnThrAerAlaPhe 1596
Db      5170 TTTTGTGACTAGCGCCCTGCAAGAACAGTGGCTTCTGCTGAGACCGCTGGGCGACCTTC 5229
Qy      1597 SerCysGluCysProLeuGlyPheGlyGlyLysSerCysAlaGlnMetAlaAerPro 1616
Db      5230 ACGTGCAGCTGCGCTGTGGGGCTTCGGCGGCAAGACTGTCACTTATGAGCCCATGCC 5289
Qy      1617 GlnHisPheLeuGlySerSerLeuValAlaThrHis---GlyLeuSerLeuProIleSer 1635
Db      5290 CACCATTCCTCGGCACCGCACACTGAGCTGGAACCTTGGAGTGAACATGCTGTGCT 5349
Qy      1636 GlnProTyTyLeuSerLeuMetPheArgThrArgGlnAlaAerProValLeuLeuGln 1655
Db      5350 GTCGCAATGTAAGTCTGGGCTGGCAATTCGACACCGGCGACGACGAGGGGTCTTAATGCA 5409
Qy      1656 AlaIleThrArgGlyArgSerThrIleThrLeuGlnLeuArgGlyGlyHisValMetLeu 1675
Db      5410 GTGACGCTGGGCGCACAGACAGCTGCTTGGCGAGTGAATGCGGGGTAACTGTCTGTG 5469
Qy      1676 SerVal---GlnGlyThrGlyLeuGlnAlaSerSerLeuArgLeuGluProGlyArgAla 1694
Db      5470 ACGATGACAGAGGAGTCCGGGCG---GTGCTTCCCATCTCTTGTGACAGAGTGACTGTC 5526
Qy      1695 AsnAerGlyAerTyThrHisAlaGlnLeuAlaLeu-----GlyAlaSerGlyPro 1712
Db      5527 AGTATGCGCGGTGACAGATCTGCGGTGAGGTGGAGGAGGAGAACAGAGTGGCGGGGG 5586
Qy      1713 GlyHisAlaIleLeu-----SerPheAerTyGlyGlnGlnArgAlaGlnGlyAsnLeu 1730
Db      5587 GGCACACATGTCCTTATGCTCACTGACCTTGAAGCTTTCAGAGACACATGCGGGTGG 5646
Qy      1731 GlyProArgLeuHisGlyLeuHisLeuSerAerHisIleThrValGlyGlyIle---ProGly 1749
Db      5647 GGAATGAGCTTGCAGAGGCTGAAGTAAAGAGCTCCACGTGGAGGGGCTGGCCCCGGGCG 5706
Qy      1750 ProAlaGlyValAlaArgGlyPheArgGlyCysLeuGlnGlyValArgValSerAer 1769
Db      5707 AGTGACAGAGAGGCTCCAGAGGTGTGGTGGCTGCACAG----- 5748
Qy      1770 ThrProGlnGlyValAerSerLeuAerProSerHisGlyGlyLysThrIleAsnValGln 1789
Db      5749 -----CAACCGGAGT-----GAAATCGGAGCT 5769
Qy      1790 GlyCysSerLeuProAerProCysAerSerAerProCysProAlaAerSerTyCysSer 1809
Db      5770 GCGTGTGTGTGACCAAGCGCTGTGCTTGGGCGCTTCCCACTGCACGACAGATGCGGG 5829
Qy      1810 AsnAerTyAerSerTySerCysSerCysAerProGlyTyTyTyGlyAspAsnCysThr 1829
Db      5830 GACCTCTGACAGACTTTTCTTGACACCTGCAGGCAAGTTACTAGCGGCCAGCGTGTGG 5889
Qy      1830 AsnValCysAerLysAerProCysGlnHisGlnSerValCysThrArgLysProSerAla 1849
Db      5890 GATGCTGCTCTCTGAACCCCTGTGCAGAACAGAGATCATGCGCGACCTGCGAGAGGCC 5949

```

QY	1850	ProHhAgLyTThrCyGgLuCyPProbAsnTyTLeuGlyProTyTCyGgLuThrArg	1863
Db	5950	CCCAATGGCTATACTGTGAATCTGTGGGTGGCTATTTCGGGACCACTGTGAACAAGG	6009
QY	1870	ILeArGInProCySPProArgGLYTrPrThrGlnHisProThrCyGgLYProCyAAsnCyS	1889
Db	6010	ATGAACCAAGAGCTTGAATCCCACTGCAACAAGAAATGGGCAAGTCACTGAAG	6069
QY	1890	AspValSerLySgLYPheAspProAspCyAAsnLyThrSerGlyLuCyAHisCyAlys	1909
Db	6070	GATGTTCAACAAGGTTTGATTCCTCACTGCAACAAGAAATGGGCAAGTCACTGAAG	6129
QY	1910	GluAAsnHisTyrThArgProProGlySerProThrCyAAsnLeuCyAAspCysTyProThr	1929
Db	6130	GAGTTCACATAACCGACCGGGGAGAGTACTTGGCTCCCAATGATGACTGTATCCCTGTGG	6189
QY	1930	GlySerLeuSerArgValCyAAspProGluAAspGlnGlnCySPProCyAAspProGlyVal	1949
Db	6190	GGCTTCACCTGGCTCATGTGACACCCCAAGCGGGCAAGTCCCTGTGGCTCCCAAGACC	6249
QY	1950	ILeGlyArgGlnCyAAspArgCyAAspAsnProPheAlaGluValThrThrAsnGlyCyS	1969
Db	6250	CTTGGCCGCCAGTGCACAGCTGTGACAGTCCCTTCGCAAGATGACAGCCAGCGGCTGC	6309
QY	1970	GluValAsnTyrThAspSerCyAAspArgAlaIleGlyAlaGlyIleTrpTrpProArgThr	1989
Db	6310	CGGGGCTCTATGAAGCTGCGCTTGAATGCTTGAATGCTGTGTGTGTGGCCCAAGACA	6369
QY	1990	ArgPheGlyLeuProAlaAlaAlaProCySPProLySgLYSerPhe-----	2004
Db	6370	AAAGTTGGCGTCTGGCCCAAGTCCCTGTCCCGGGGGGGCCCTGGAGATTGGGGGTGCA	6429
QY	2005	GlyThrAlaValArgHisCyAAspGlnHisValArgLYTrpLeuProProAsnLeuPheAsn	2024
Db	6430	GATGCTGCTGTGGCGGCTGTGTGATGAAGCCCAAGGTTGGCTGAAGCCCACTTTCAC	6489
QY	2025	CysThrSerSerIleThrPheSerGlnLeuLySgLYPheAlaGluValGlnGlnArgGln	2044
Db	6490	TGTACTCTCCCTGCGCTTTCGAAAGTCAAGTCTGCTGTAAGGCTTGAAGCTGAACAAG	6549
QY	2045	SerGlyLeuAspSerGlyArgSerGlnIleLeuAlaLeuLeuArgArgAsnAlaThrGln	2064
Db	6550	AACGACACTGATACATGAGAGGCCAAGAAGCTGGCTACGCGGCTACGGGAAGTGACTGC	6609
QY	2065	HisThrAlaGlyTyrPheGlySerAspValLyValAlaTyrGlnLeuAlaThrArgLeu	2084
Db	6610	CACACTGACCACTATTTTAGCCCAAGATGTTCAGACTCACTGCCCGCTGCTGACCACTGC	6669
QY	2085	LeuAlaHisGlySerThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPhe	2104
Db	6670	CTGGCTCTCGAAGACATGAGAGGGGCTTCGGGCTGACAGCCACAAGAAATGCCCACTTC	6729
QY	2105	ThrGluAAsnLeuLeuArgValGlySerAlaLeuLeuAspThrAlaAsnLyArgHisTyrP	2124
Db	6730	AATGAAGAAATGCTGTGGGCGGCTCTGCACTGTGGCCCAAGACAGGGGAATGTGGG	6789
QY	2125	GluLeuIle---GlnGlnThrGlnGly-----GlyThrAlaTrpLeuLeuGlnHis	2140
Db	6790	GCGGCGCTGGGGCACGCGGCGCTTCGGGGGCTCCCAAGGACAGCGCGGACCTGTGAGGAC	6849
QY	2141	TyrGluAlaTyrAlaSerAlaLeuAlaGlnAsnMetArgHisThrTyLeuSerProPhe	2166
Db	6850	CTGGAGAGATATGACGCCCACTCCGCAAGAAATATGAATCACTCACTCAATATCCATG	6909
QY	2161	ThrIleValThrProAsnIleValIleSerValValArgLeuAspLySgLYAsnPhe---	2179
Db	6910	GGGCTGTGTGACGCTTAATATCATGCTCAAGACTGAACGAGATGAAGCAACCCAGTTCTCC	6969
QY	2180	---AlaGlyAlaLyLeuProArgTyrGlnAla-----LeuArgGlyGlnGlnProPro	2196
Db	6970	CGGGGGGCGCGCTGACTCTTCGTAACAAGCAACTCTTTCAGAGGCCAAGATGCGCTGC	7029

QY	2197	AspLeuGluThrThrValIleLeuProGluSerValPheArgGluThrProProValVal	2216
Db	7030	GATCTCTACACCCATGAGTGCCTGCTCCACGATCCCAAGGCATCCCATCTGAAGTT	7089
QY	2217	ArgProAlaGlyProGlyGluAlaGlnGluProGluGluAlaArgArgGlnArgArg	2236
Db	7090	CTGGCCACAGACAGACAGATAGAAACTCCACCAACCTCAAGTGTGATGCCCAACGCC	7149
QY	2237	HisProGluLeuSerGlnGlyGluAlaValAlaSerValIleIleTyPArgThrLeuAla	2256
Db	7150	CCGCACAGCAGACAGCCTGGGATCTCCATATTCATCTCCCTGTTACCGGACCTTAGG	7209
QY	2257	GlyLeuLeuProHisAsnTyPAspProAspArgArgSerLeuArgValProValArgPro	2276
Db	7210	GGAATGCTCTCTGCCCAAGTTCACAGACAGAACCCAGAGTGCAGAGCTTCTCAAGACCC	7265
QY	2277	IleIleAsnThrProValValSerIleSerValHisAspArgGluLeuLeuProArg	2296
Db	7270	GTCATGAACTCCCGGTGTCTACAGGTGAGTGTGTTCACGCAAGCACTTCTTAAGGGGA	7329
QY	2297	AlaLeuAspArgProValThrValGlnPheArgLeuLeuGluThrGluAlaArgThrLys	2316
Db	7330	ATCTCGAAGTCCCCATACGCTAGAGTTCGCTGCTACAGACAGCGAATCGGAGCAG	7389
QY	2317	ProIleCysValPheTrpAsnHisSerIleLeuValSerGlyThrGlyTyPTrpSerAla	2336
Db	7390	GGATCTGTGTGCAGTGGACCCACCTGGCTGGGAGACAGACATGGTGTGTGACAGCA	7449
QY	2337	ArgGlyCysGluValValPheArgAsnGluSerHisValSerCysGlnCysAsnHisMet	2356
Db	7450	CGGACATCGCAGCTGGTGACAGAGAAATGGGTCCACGCAACGATGCTCGACGCGGACA	7509
QY	2357	ThrSerPheAlaValLeuMetAspValSerArgArgGlu-----AsnGlyGluIleLeu	2374
Db	7510	GGGACCTTTGGGGTCTCTATGAGTCCCTCTCCCTGAGAGCTGAGAGCGGACCTGGAG	7565
QY	2375	ProLeuTyPThrLeuThrTyPValAlaLeuGlyValThrLeuAlaAlaLeuLeuLeuThr	2394
Db	7570	CTGCTGGGTGTGTTACCCACAGTGTGTGTGGCTGTGTCTGTGTGCTGGCTGTGTGACT	7629
QY	2395	PhePhePheLeuThrLeuLeuArgIleLeuArgSerAsnGlnHisGlyIleArgArgAsn	2414
Db	7630	GCAGCCATCTCTGACAGCTGTGCGAGCTTCAGAGCTTCAGATTCATGTGCGTGGATTCAGCCAT	7689
QY	2415	LeuThrAlaAlaLeuGlyLeuAlaGlnLeuValPheLeuLeuGlyTyrIleAsnGlnAlaAsp	2434
Db	7690	GTGGAGACCGCTCGGGGTGGAGAGCTCTTCCGTGCTGGGATTCACAGACCCAC	7749
QY	2435	LeuProPheAlaCysThrValIleAlaIleLeuLeuHisPheLeuTyPLeuCysThrPhe	2454
Db	7750	AATCAGCTGTGTGCATCAGTGCAGTGCAGCATCTCTGCACTATCTTCTTCAGACCTTC	7809
QY	2455	SerTrpAlaLeuLeuGlnAlaLeuHisIleLeuTyPArgAlaLeuThrGluValArgAspVal	2474
Db	7810	GCGTGTGCTCTTCGTGCAGGCGGTGCACCTCTACCGCATGACAGTGTGAGCCACGCAAGTG	7869
QY	2475	AsnThrGlyProMetLeuArgPheTyPTrpMetLeuGlyTyPArgValProAlaPheIleThr	2494
Db	7870	GACCGCGCGCCAGACGCTCTTACACAGCCCTGGGCTGGGGCTGCCCTGCTGTGCTGTG	7929
QY	2495	GlyLeuAlaValAlaGlyLeuAspProGlnGlyTyPArgIleAsnProAspPheCysTyPLeuSer	2514
Db	7930	GCGCTTGCTGTGGGCTGTGACCTGTGAGGCTATGTGGAACTTGACTTGTGTGATCTCA	7989
QY	2515	IleTyPAspThrLeuIleTyPTrpSerPheAlaGlyProValAlaPheAlaValSerMetSer	2534
Db	7990	GTCCACGAGCCCTCATCTGTGAGCTTTGTGGCCCTTGTGCTGTGATCATGTATGATGAC	8049
QY	2535	ValPheLeuTyPTrpIleLeuAlaAlaArgAlaSerCysAla---AlaGlnArgGlnGlyPhe	2553
Db	8050	GGGACCATGTTTCTCTGCTGCCCGGACATCTCTGCTCACAGGCGAGGGGAAGCC--	8106
QY	2554	GluTyPArgGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeuLeuSer	2573



Db	543	GTGCACATATACATCATCATGATGGCAATGCTCGGAGGACAGTTTATCTGGATGCCAGACT	602
Qy	454	GLVLALEuAaPbVAlValSerProleuAaPbYrGIuThrThrIlybGIuTYrThrLeuArg	473
Db	603	GGAGCTCTGGATAGTGGTGAAGCCCTTCACTATAGAGAACCAAGATACACCTTACAGG	662
Qy	474	VALArgLAglAaPbGLyGLYArGPbProleuSerAaenValSergLYLeuValThrVal	493
Db	663	GTGGAGACAGAGATGGTGGCCGTGCCCACTCTTATGTCTCTGGCTTGGTGAAGTA	722
Qy	494	GLInValLeuAaPbLIeAaAaPbAaAlaProLIePheValSerThrProPheGLInAlaThr	513
Db	723	CAGGTCTCGATATCAACAGAACATGCCCCCACTCTTCGACAGACCCCTTCCAGGCTACT	782
Qy	514	ValLeuGIuSerValProLeuGIYTYrLeuValLeuHISValGLInAlaLIeAaPbAaSP	533
Db	783	GTCTCGAGAGAGCTGCCCTTAGGCTACTGTCTTCATATGTCACAGGCTATGAGCGCTAT	842
Qy	534	AlaGLYbAaPbAaAlaAaArgLeuGIuTYrArgLeuAlaGLYValGLYHISaSPbPheProHe	553
Db	843	GCTGGTGAACAATGCCCGCTGGAAATACCGGCTTCGTGGGGTGGGACATGACTTCCCTTC	902
Qy	554	ThrLIeAaAaenGLYThrGLYThrProLeSerValAlaAlaGLYLeuAaPbArgGLYGLYVal	573
Db	903	ACCATCAACAATGGACAGGCGCTGATCTCTGTGGCTGTCAACTGAACTGAACTGGAGAAATTT	962
Qy	574	AaPbPheTYrSerPheGLYValGLYAlaArgAaPbHISGLYThrProAlaLeuThrAlaSer	593
Db	963	GATTTCTTACAGCTTTGGGTGGTGAAGCTCGAACATGCACTCGACACTCATGCTCTCG	1022
Qy	594	AlaSerValSerValThrValLeuAaPbValAaAaPbAaAaenProThrPheThrGLInPro	613
Db	1023	GCCAGTGTCAAGCGTGAAGCTGCTCGATGTCAACGACAAATTCACACTTATACCAACACA	1082
Qy	614	GLYUryThrValAArgLeuAaenGLYbAaPbAlaAlaValGLYThrSerValValThrValSer	633
Db	1083	GAGTACACAGTGCGGCTCAATGAAGATCAAGCTGTGGGACACAGGCTGTGAAGGTGCTCA	1142
Qy	634	AlaValAaPbAaPbAaAlaHISerValLIeThrTYrGLInLIeThrSergLYAaAaThrArg	653
Db	1143	GCTGTGGACCGTGATGCTCATATGTCTATCACTTACAGAACATCAAGTGGCAATTACTCGA	1202
Qy	654	AaAaPbPbPheSeriLIeThrSergLIuSeriGLYGLYLeuValSerLeuAlaLeuProleu	673
Db	1203	AACGAGCTTCTCATATCACAGGCCAAAGTGTGGTGGAGCTGTGATTCCTGTCCCTGCACTG	1262
Qy	674	AaPYrLYLeuGIuArgGLInTYrValLeuAlaValThrAlaSerAaPbGLYThrArgGLIn	693
Db	1263	GACTACAAACTTGAAGCGGCGAGATGTGTGGCTGTTAACGGCTCGAATGGCACTCGGAG	1322
Qy	694	AaPbThrAlaGLInLIeValAlaAaenValThrAaPbAlaAaenThrHISArgProValPheGLIn	713
Db	1323	GACACGGACAGATTTGGTGTAAATGTCAACGACGCAACACCACATCGTCTGTTCAG	1382
Qy	714	SeSerHISerTYrThrValAaenValAaenGLYbAaPbArgProAlaGLYThrThrValValLeu	733
Db	1383	AGCTTCCCACTTACAGTAATCTTATATAGAGACGGCCGGGAGGACACACGGTGTGTGTG	1442
Qy	734	LIeSerAlaThrAaPbLIuAaPbThrGLYGLYbAaenAlaArgLIeThrTYrPheMetGLYbAaP	753
Db	1443	ATTCAGCGCCAGAGTGAAGACACAGGTGAAGATGCCCAATCACTTATCTATGGAGAGAC	1502
Qy	754	SeLIePbProGLInPheArgLIeAaPbAlaAaPbThrGLYAlaValThrThrGLInAlaGLYLeu	773
Db	1503	AGCATCCCCCAAGTTCCGATGCAATGCAACACGGGGGCTGTCCACCACCCAAGCTGAAGCTG	1562
Qy	774	AaPYrGLYbAaPbGLInValSerTYrThrLeuAlaLIeThrAlaArgAaPbAaenGLYLIePro	793
Db	1563	GACTACAGAAACCAAGTGTCTTACACCTTGGCCATTTACTGTCCGGGACAAATGGCAATTCCC	1622
Qy	794	GLInLYSerAaPbThrThrTYrLeuGLYLIeLeuValAaAaPbValAaAaPbAaAlaPro	813
Db	1623	CAGAAAGTCCGACACCACTTACTGTGAAGATCTCGTGGAAACAGGTGAATGAACAATGCCCTT	1682

QY	814	GLPheLeuArGAspSerTYrGInGlySerValLYrGIuLAspValProPheThSer	833
Db	1683	CAgTTCCTGGAGACTCTCAACAGGGCAGTGTATGAGATGTGCCACCTTCACATGAC	1742
QY	834	ValLeuGInLIeSerAlaThrAspArGAspSerGlyLeuLanGlyArGValPheYrThr	853
Db	1743	GTCTGTGAGATCTCAGCCACTGATCGATTCGAGACTTAATGACAGGGCTTCTTCAACC	1807
QY	854	PheGInGlyLYAspAspGlyAspGlyAspPheLIeValGIuSerThrSerGlyLIeVal	873
Db	1803	TTCCAGAGAGAGCCAGCATGTGAGAGCGGTGACTTTATTTGTGTAGTCCACGTCAAGCATCGTG	1866
QY	874	ArgThrLeuArGArGLeuAspArGGLuAsnValAlaGInLYrValLIeArGAlaLYrAla	893
Db	1863	CGAACGCTTACGAGGCTGTGATTCAGAGAAAGTGGCCCGCATGTGCTTGCGGGGCAATATGCA	1922
QY	894	ValAspLYrGlyMetProProAlaArgThrProMetGIuValThrValThrValLeuAsp	913
Db	1923	GTGACACAGGGGAGATGCCCCAGCGCCGACACTGTGAGAAAGTGAACATCATCTGTGTGGAT	1982
QY	914	ValAsnAspAsnProProValAPheGInGlyAspGluPheAspValAPheValGInGluAsn	933
Db	1983	GTGAATACAAATCCCTCTGTCTTTGACAGAGATGATTTGATGTCTTTGTGGAAAGAGAAC	2042
QY	934	SerProLIeGlyLeuAlaValAlaArgValThrAlaThrAspProAspGluGlyThrAsn	953
Db	2043	AGGCCCCATTGGGCTAGCGCGGTGGCCCGGTGTACAGCCACTGACCCCGATGAAGGACCACT	2102
QY	954	AlaGInLIeMetYrGInLIeValGInGlyAsnLIeProGluValAPheGInLeuAspLIe	973
Db	2103	GCCCAAGTTATGTACCAAGATTGTGTGAGGGGCAACATCCCTGAGGTCTTCCAGCTGGACATC	2162
QY	974	PheSerGlyGlyLeuThrAlaLeuValAspLeuAspTYrGlyAspArgProGluTYrVal	993
Db	2163	TTCTCCGGGGAGGTGACAGCCCTGTGTAGACTTGTACATCAAGAGACCGGCTCGAATACGTC	2222
QY	994	LeuValLIeGInAlaThrSerAlaProLeuValSerArgAlaThrValAlaValArgLeu	1013
Db	2223	CTGGTATCTACAGGCGACAGTCAGGTCCCTCTGTGTAGCGGGGTACATGCTCAAGTCCGCGCTC	2288
QY	1014	LeuAspArGAsnAspAsnProProValLeuGInGlyAsnPheGluLIeLeu	1029
Db	2283	CTTGAACCGCATATGACCAACCAACAGAGCTGGGCAACTTTGAGATCTT	2330

Search completed: February 14, 2004, 11:45:48  
Job time : 3172 secs





414 -VEIDARSGVTRAVVDREBAAYQLVEANDQGNPPLSASTVHI VVEDENDT 472  
QY 398 QFSEKRYVQVREDVTPGAEVLRTASDRKGSNAVHVSIMGNARGOFTYLDQATGD 457  
DB 473 QFSEKRYVQVREDVAVNTAVLAVQATDRDQGNAAIHVSIVSGNKGQYHLSJSGSLD 532  
QY 458 VSPRLDYETTKETLAVRADGGRPPLSNVSGVTVQVLDINDNNAIFVSTPQATVLS 517  
DB 533 VNPDLFEALREYTLTKAQDGRPPLINSGLVSVQVLDVNDNABIPVSSPQAAVLN 592  
QY 518 VPLGYLVHQAIDADAGNARLEYRLAGVH-----DFFPTINGTG 560  
DB 593 VPLGSHVLAHQADADAGNARLEYRLVDTASTIVGSSVDSKNPASPDPFQIHNSSG 652  
QY 561 MISVAALDREEDVFPYSGVYABDHGTPTALTASASVTVLDVNDNPFPTQPEYTVRLN 620  
DB 653 WITVCALDREEDVEHVSFGVEAVDHGSPAMSSASVSTVLDVNDNPMFTQVVELRLN 712  
QY 621 EDAAVGTSVTVANVDRDASVTTQITSGTNRBSITQSGGLVSLALPDYLERQ 680  
DB 713 EDAAVGSSTVTLARBDANSVTTQITGNTNRFPALSSQSGGLITLALPDYQERQ 772  
QY 681 YVLAATASDGTRODTAQIVNVNTDANTHRPVFQSSHVTVNVNDRPAGTVVLISATDE 740  
DB 773 YVLAATASDGTSHTAQVFINVTDANTHRPVFQSHVTVSVDREPVGISATISATDE 832  
QY 741 TGNARITVPMEDSIPQFRIDADTGAVTQABLDEDOVSYTLATARDNGIPQKSDTY 800  
DB 833 TGNARITVYLEDPVFPQFRIDPTGTITVTELDYEDQAAVTLATARDNGIPQKSDTY 892  
QY 801 LELIUNDVNDNAQOFLADSVQSVYEDVFPFTSVLQISATDRSGLANGRFYTFQCGDNG 860  
DB 893 LELIUNDVNDNARFLADFYQSVFEDAPSTSVLQVSAIDRSNGRGLLYTFQCGDNG 952  
QY 861 DGDPIVESTSGIVRTLRRLRENVAAQVLAAYAVDKAMP--PARTPMEVTVTVLVNDNPP 919  
DB 953 DGDPIFEPISGVRTQGRDLRENVAAVYNMALAVDKSPMLBASVGIQVSTVDINDNP 1012  
QY 920 VPEODEPVEENSPITGLAAVANVATADPDGNTAQIMYOIVEGNIPVEFQDILFSGELT 979  
DB 1013 VPEKDELELFEVBENSFVGSVVAARIRANDPDEGPAQIYQIVEGANVEVQDILLSGDLR 1072  
QY 980 ALVDLDYEDREPEVTVLOATSAPLVSATVHVLBRNDNPPILGNPEILLFNMYNNRBS 1039  
DB 1073 ALVELDREVRDMLVQATSAPLVSATVHILRNDNPPILGNPEILLFNMYNNRBS 1132  
QY 1040 SPPGCAIGRPADPDIDSLTYSFPERGNEISVLINASTGELKSRALDNNRPLAINS 1099  
DB 1133 SPPSGVIGRIIPADPDIDSLNTTFLQGNELSLILLDPATGELQSRDLNRRPLAINS 1192  
QY 1100 VLVSQVHSTAAOCALRVTTITDEMLTHSITLLEDMSPERPLSPILGLFTQAVATLAT 1159  
DB 1193 VVSVDGHSVTAALCTLRVTIITDDMLTNSITVRELENNSSQEKPLSLSTLFEVGAATVLS 1252  
QY 1160 PRPHVVVFNORBDAPCGHILNVSIVQPPGCGGPPPLPSEDOERLXTNRSLTAT 1219  
DB 1253 TKODIFEFNQNBDV--SSNITLNTFSALLPGGRG--RFPSEBDQEQYINRLLTIT 1309  
QY 1220 SAORVLPFDNITCLREPCENYMRCSVTLRFDSPAPFASSVLFRPIHPGRLCRCPG 1279  
DB 1310 SAORVLPFDNITCLREPCENYMRCSVTLRFDSPAPFASSVLFRPIHPGRLCRCPG 1369  
QY 1280 FTGQYCEVBYDLCSRPCGPHGRCSRREGYTCICRQDYGTHGECEVASRSGRCTPVYCN 1339  
DB 1370 FTGQYCEVBYDLCSRPCGPHGRCSRREGYTCICRQDYGTHGECEVASRSGRCTPVYCN 1429  
QY 1340 GGTGVNLLVGFCKDCSPDEKPYCOVTRSPFAPHSPIFRGLRORFHTLALSPATKE 1399  
DB 1430 GGTGVNLLVGFCKDCSPDEKPYCOVTRSPFAPHSPIFRGLRORFHTLALSPATKE 1489  
QY 1400 RDGLLLYNGRNEKRDVALEVLEQVQVLTFSAGSSTTVSPVPGVSDGOMHTVQAKY 1459  
DB 1490 RNALLLYNGRNEKRDVALEVLEQVQVLTFSAGSSTTVTPQVPGVSDGOMHTVQAKY 1549

QY 1460 YNKPILLGQITGLPEQGBEOKVAVTVYDGDVVALRFGSVLGNYSCAAQSTQSGSKSLDL 1519  
DB 1550 YNKPINIGHILPHGSGSKVAVTVYDGDVVALRFGSVLGNYSCAAQSTQSGSKSLDL 1609  
QY 1520 TGPILLGGVPLDPSEFPVRMQPVGCNRNLOVDSHIMADPIANGVIRPGPACQVND 1579  
DB 1610 TGPILLGGVPLDPSEFPVRMQPVGCNRNLOVDSHIMADPIANGVIRPGPACQVND 1669  
QY 1580 SNTCANGTQVQMAFSCCEPLGFGKSCAQBEMANPHFLGSSLVAMHGLSLPISQPMY 1639  
DB 1670 GTSQNGGTQVQMAFSCCEPLGFGKSCAQBEMANPHFLGSSLVAMHGLSLPISQPMY 1729  
QY 1640 ISLMFRTQADGVLLQAITRGASTTTQLREGVNLVSEGTGLQASSLRLBERANDGDW 1699  
DB 1730 ISLMFRTQADGVLLQAITRGASTTTQLREGVNLVSEGTGLQASSLRLBERANDGDW 1789  
QY 1700 HHAQALALASGPRH-----ALISFDYQQAABGRLGLHLSNTTVGGIIPBPAGV 1754  
DB 1790 HHAQALALASGPRH-----ALISFDYQQAABGRLGLHLSNTTVGGIIPBPAGV 1848  
QY 1755 ARGFRCLOGVVSDTPGVSINLPSHGESINVEQCSLPPPCDGNPCPANSYCSNDMS 1814  
DB 1849 RHGFRCLOGVVSDTPGVSINLPSHGESINVEQCSLPPPCDGNPCPANSYCSNDMS 1908  
QY 1815 YSCSDPGEYDGNCTNVCLNPECHQSVCTKPSAPHGYCECPNVLAPYCEIRIDPC 1874  
DB 1909 YSCSDPGEYDGNCTNVCLNPECHQSVCTKPSAPHGYCECPNVLAPYCEIRIDPC 1968  
QY 1875 PRGMWGHPTCPGNCNDVSKGPDCKNTSGRCHCKENHRRPGSPTCLLDCYCPGSLR 1934  
DB 1969 PRGMWGHPTCPGNCNDVSKGPDCKNTSGRCHCKENHRRPGSPTCLLDCYCPGSLR 2028  
QY 1935 VCDPBDGQCPCKPVGIGRQCDRCNDPFAVTTNGCEVNVYDSCPRALTEAGIWMPTFRGLP 1994  
DB 2028 VCDPBDGQCPCKPVGIGRQCDRCNDPFAVTTNGCEVNVYDSCPRALTEAGIWMPTFRGLP 2088  
QY 1995 AAAPCPKSPGTVAVHCDHGRGMLPPLNCTSTISFBLKGFABRLQNNESGLDSGRSQ 2054  
DB 2088 AAAPCPKSPGTVAVHCDHGRGMLPPLNCTSTISFBLKGFABRLQNNESGLDSGRSQ 2148  
QY 2055 LALLRNATHTAGVFGSDVYKAYOLATRLAHESHTORFGLSATODVHFTENLRVSA 2114  
DB 2149 LALLRNATHTAGVFGSDVYKAYOLATRLAHESHTORFGLSATODVHFTENLRVSA 2208  
QY 2115 LLDTANKRMWELIQQTEGTAWLQHYEAYASALAQNNRHTYLSPTTIVPNIIVISVRL 2174  
DB 2209 LLDTANKRMWELIQQTEGTAWLQHYEAYASALAQNNRHTYLSPTTIVPNIIVISVRL 2268  
QY 2175 DKGNFAGAKLPRYBALRGEOPDLETTYILPESV-----RETPPYVR-----PAGP 2221  
DB 2269 DKGNFAGAKLPRYBALRGEOPDLETTYILPESV-----RETPPYVR-----PAGP 2328  
QY 2222 GSAOPEBELARORHPELSQGEBAVASIYRTLAGLLPHNYDPKRSIRVPRPINTP 2281  
DB 2329 GSAOPEBELARORHPELSQGEBAVASIYRTLAGLLPHNYDPKRSIRVPRPINTP 2388  
QY 2282 VVSIVHDEBELLPALDKPVTVQRLLETERTKPCIVFNNHSLIVSGTGSMSARCEV 2341  
DB 2389 VVSIVHDEBELLPALDKPVTVQRLLETERTKPCIVFNNHSLIVSGTGSMSARCEV 2448  
QY 2342 VFRNESHVSQCCNHTSPAVLMDVSRRENGELLPKLTLYVALGYTLAALLTFPFLTL 2401  
DB 2449 VFRNESHVSQCCNHTSPAVLMDVSRRENGELLPKLTLYVALGYTLAALLTFPFLTL 2508  
QY 2402 RILRSNOGIRBANTRALAQVFLGNOADLPACTVIALIHLFTVLCFSSVALLA 2461  
DB 2509 RILRSNOGIRBANTRALAQVFLGNOADLPACTVIALIHLFTVLCFSSVALLA 2568  
QY 2462 LHLRYALTEVRDVTGPMRFYTMGLGWPAFTYGLAVGLDPRGYGNPDFCWLSTYDTLW 2521  
DB 2569 LHLRYALTEVRDVTGPMRFYTMGLGWPAFTYGLAVGLDPRGYGNPDFCWLSTYDTLW 2628

QY 2522 SPAGPAPAVSVMFLYITIAARASCAAQOGFEKGPVSGLOPFAVILLLSATWILL 2581  
 DB 2629 SPAGPAGVYIIINTVIFVLIAKVSQCKHRYERKGVSMILRTAFILLLVATWILLGL 2688  
 QY 2582 SVASDILLFHYLPATCNICIGPFILSYVVLSEKVRKAL-LACSKRSPDPLTTKST- 2639  
 DB 2689 AVNSDILLFHYLPATCNICIGPFILSYVVLSEKVRKAL-LACSKRSPDPLTTKST- 2748  
 QY 2640 LTRSYNCPSPYADG-RLYOPYDSDAGSLHSTSRSGKSPSYLPFLREESALNPGQGP 2697  
 DB 2749 LTRSLNCNNYSEGPMILRTALGESSTASLDSTTRDGVQ-----KLSTSSGPAG 2798  
 QY 2698 GLQDPSGLP-EGQDQHPDPTSDSDLSLEDQSGSYASTHSDSEEEEEEPAAP 2756  
 DB 2799 NHEBPTSTPLPNSKKAHGDSDSDRLSL-DEHSSSYASTHSDSEEDGGEALDK- 2853  
 QY 2757 GEGMDSLGPGARLPLHSTPK-DGPGPGKAPWBGD-FGTAR- 2800  
 DB 2854 ----WNPAGGA-----HSTPADLALNHPVAGWDESLAGSDSEELDTEPLKVTY 2903  
 QY 2801 -----SSGNAPBERLRENGDALSRGSLGPLPGSSAP--HKGLKKCL-PTIS 2848  
 DB 2904 SVELHQAQGNHCGDRPSDESGVLAK-----PVAVLSSQPGQRKGLKNTYPPPLP 2958  
 QY 2849 EK--SSLPLPLEOCTGSSRG-SASGSRGPP-----RPPRQSLQEQINWMP 2898  
 DB 2959 EGPLKRLRKLKADCEQSPSTSSISLGGDGVHATDVTITKTPRPRERETLNGV- 3015  
 QY 2899 AMSIKAGTVDEDSGSE 2915  
 DB 3016 AMNVRTGSAQANGDSSE 3032

## RESULT 2

T20968  
 Hypothetical protein F15B9.7 - Caenorhabditis elegans  
 C.Species: Caenorhabditis elegans  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C.Accession: T20968; T26278  
 R.Pericy, C.  
 submitted to the EMBL Data Library, August 1996  
 A.Reference number: Z19351  
 A.Accession: T20968  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-2610 <W1>  
 A.Cross-references: EMBL:Z78013; PIDN:CAB01427.1; GSPDB:GN00023; CESP:F15B9.7  
 A.Experimental source: clone F15B9  
 R.Baynes, C.  
 submitted to the EMBL Data Library, August 1996  
 A.Reference number: Z20187  
 A.Accession: T26278  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-2610 <W1>  
 A.Cross-references: EMBL:Z78018, PIDN:CAB01449.1; GSPDB:GN00023; CESP:F15B9.7  
 A.Experimental source: clone W07G4  
 C.Genetics:  
 A.Gene: CESP:F15B9.7  
 A.Map position: 5  
 A.Introns: 169/2; 217/3; 239/3; 284/3; 515/2; 550/3; 738/3; 810/1; 838/3; 890/2; 977/1;

Query Match 16.0%; Score 2491; DB 2; Length 2610;  
 Best Local Similarity 27.4%; Pred. No. 8.2e-119;  
 Matches 768; Conservative 438; Mismatches 1205; Indels 394; Gaps 89;  
 QY 51 CAPMGILCSSASNLMLYSR--CRDAGRLGHLVPHHDGLRWCPESBAHPLPAP 108  
 DB 39 CRCA--VSSNSVILPASRPPCLHPGQPII-HMPDLSDNA-----CPVPGLE 86  
 QY 109 GCPWSCRLLGIGHLSPQGLTLPEEHPC-LKAP-----RLRQSCGLAQAPGL 156  
 DB 87 SVHSGISL-----LEGSG-LILTKERICFPDGPIDPHDYCDGCLYNSKRIHGIS 140

QY 157 RAGERSPEESLGGRRKRWNTAPQFPSPSYQATVENOPAGTPVASLRAIDDEBAGRL 216  
 DB 141 KKKLETRTKRARRRRNDANAVHQQEKYVLEBEDPIETIIISVKA---SHASQGL 197  
 QY 217 EYTDALFDSRNOFPISLDPVTGAVTTABEIDRETKSTHVFVYTAODHMPRSALATLT 276  
 DB 198 YSMAVAPQDSRQNFPTLDWSGEIRLAKSMOREVLDKILLYATVERVDPITISATTV 257  
 QY 277 ILVTDNDHPVEEQGKESRENLVEYEVLTATRDGDAPRANILYRLREGSGSP 336  
 DB 258 VAVLDVQDSPLFEKOSTFGBIREDAPIGTLYSVFADLDGEGEIBLYSAGEGNG- 314  
 QY 337 SEVFEIDPRSGVTRRGVDEEVEESYOLTYEASQDGRDPGRSTAAVFLSEVDNDNA 396  
 DB 315 KULAIANKSGVIGTAAPLDRETTLSLILDVIAKGP--KRESTANVEIIVDVNDNA 372  
 QY 397 PPSERKRYVOVREVTGAPVLRVYASDRDKSNAVHYISIMSGNAGQVTLAQGTAL 456  
 DB 373 PVFADSYNVVTLLENITPAVATATKATDEDFGTNGKHYSWASSSGIGLTIIDYSTGEV 432  
 QY 457 DVVSPLDYETTEKYLRAADQGRPLSNVSGLVTVGLDINDNAPFVSTPPQATVE 516  
 DB 433 TIKERIDAKNS-PITAVIRAKDGAQALSTVPL-TINVIDINDAPFLIAQKMITLE 490  
 QY 517 SVPLGLVLAQADADADGNARLEYRLAGVGHDPFTINGGWI SVAABLDREEDVPY 576  
 DB 491 NVALIGEVGRVYALDEDSGPNGLIKYMEG---SDFIIDEBSGLIKTKLDRRTARY 547  
 QY 577 SPGEARDHGTPALTAGASVSVTVLDVNDNPTFPQETRYRLNDAAVGTSVTVSAVD 636  
 DB 548 SLKVTARMDGTPSLNTSTTIAVVLINDNAPFPDKKEYNVLTISEMRGSOIITLKAVD 607  
 QY 637 RDAHSVITYQITSGNTRRFSITSGGGGLVSLPLYKLERO---VLAATYASD-GTR 692  
 DB 608 NDEDKITIRLEAD-REYFSIILIDGQGA---LSVSGELRKQDHKRAVEISATDQGL 663  
 QY 663 OPTAQIVNVTDANTHREVPFQSHYTVNVNEDRPAQTVVLIS----- 735  
 DB 664 QGRCVANVFIDVNS-APYFNDHPSVKLPEHSFGYVITLKVAFSGYFGYGLKHFFV 722  
 QY 736 -----ANDEDTGERARTTYEMEDSIPOFRIDADQATTOALEYEDOV 779  
 DB 723 SSNLKRRKIPNSAMWMAEDHGRGNARIVYSI-DSSQFRIDPDSGDSVSDILDRBREA 781  
 QY 780 SYTLAITARNDGIPQKSDTYTEILVNDVNDNAPFLASQSYVEDVPPTSVLQISA 839  
 DB 782 TFSVAVTASDASPPPLNTSTQIEVLIDINDSPQFTSSSYAATISBDIPVGTSLQVSA 841  
 QY 840 TDRSGLNGRVYTFQGGDDGDGD--FVSESTSGIVRTLRLDRENVAAQVYLRAVAVDG 897  
 DB 842 IDADIGPAGVDFYFNBSGSSPSIQFLDRTSGTLARVSKLDRGQFAVTVLPFARBG 901  
 QY 898 MPRAATPMEVTVTVLVNDNPPVFEQDFDVYVENSIGLAVARVTTDDEGNAQIM 957  
 DB 902 TFSLSAABEITLTSVDVNDNAPTFEQLSYDYLIENSSVGSITGVIVARDDEGNADIS 961  
 QY 958 YQIVEGNIPEVFOIDIFSGE-----LTAIVLDYEDRE-EVYLVQATSAPVSRATYHV 1011  
 DB 962 FRIFGALAKIFDIEBDAKQNGVRIILTRAFTDEAKANKFFELQASSGQLSTVPR 1021  
 QY 1012 RLDDNDNPVLGNFEILFNNTVNRSSSPGAGIAGRPAHDPSIDSLTYSFERGNETS 1071  
 DB 1022 HVSVDNDKPKLPKDFVILMNFVNDVQAR-----QIGFIPARDPDQNTALRYFL- 1074  
 QY 1072 LVLLNASTGELKLRALDNRPLRLAKSVYSDGSHVSTAQACARVITITBMLTHSTTL 1131  
 DB 1075 LIAEKYTKLIVQEWGRN--MDVSFRTCVSDGANTRESCRTFTHLVBEWMLSESTFL 1132  
 QY 1132 RLEDMSPERFLSPILGLFQAVATLAT-PPDHVVVFVQGDTPDAGHILNVSLSVQOP 1190  
 DB 1133 SLARTTVDDPDLVFGQFRDAMSTLSNWKPSDHIHVGKQHD-DVIYINIAITDHR 1190

```

Qy 1191 PGPGGPPLPESDQRLYLNRSLTAISAORVLPEDDNLCLREPCENTMRCVSLRFD 1250
Db 1191 VVGWRAIHELKESIKK---LEKWTLLQVEYIR-----DSSCANEPCHMAKQCTQKFV 1242
Qy 1251 SSAPFIASSSVLPFPIHVGGLRCRCPPGPT-----GDYCEYDLCYSRPPCHGRCRS 1305
Db 1243 GEMKAMHDTNFIAITLNTVTFVCEPCSGFTSSGAMBD-CDTRIDECYRGRCSNNSVCVA 1301
Qy 1306 REGGYTCLRDYGEHCEVASRSGRCSPGVCKNGTGVNLLVGGFKC-DGP-SGDPEKP 1363
Db 1302 FENTYQCECKRGWGRHCEISVHALTCVPKCMSDSC-BLDGNQMKCRCKHGEPTDE 1360
Qy 1364 YCQVTTSPFASPTFRGLRQRFHTLALSFATKEDGLLYNGRPNKHDVVALLEVIQ 1423
Db 1361 RCRLRSVSPDEGHLNVLNLDLPTQMTMKPFVSTIAHGVLVFTG--DKSDPFEVSVVD 1418
Qy 1424 EQVQVLTSSAG--ESTTIVSPFVPGVSDGQWHTTQKTYNKPFLIGQGLRQGPBEQKAV 1481
Db 1419 RVLKVQPSLGGEBKIDAKMENDVENRINDGEMHTVAALEYSNKQI-----T 1462
Qy 1482 VTVNGCPTGVALRFSGVLYANYSACAGTGGGSKS-----LDLTGPIILGGVPLD 1531
Db 1463 MSLLDDCETNELLIANT---SPNCAIRAKNLKCEPPTVPCRYLIDISNGLPFGGRPBT 1519
Qy 1532 PESFVWRQFVGCNRLQVDSRHIDMAF--IANNGTV-PGCPAKXNVCDSN--TCANG 1586
Db 1520 SKQJE---KAFGGCIISLSDVDEVDPSTIKEMHKVQVHGCGHGRDPCSTSDGQCSAT 1576
Qy 1587 GTCYNQNDAPSCCEPPLRFGGSCAQEBANRPHFGSSLVAMHGLSL----PIGQPTLSL 1642
Db 1577 SKCNRRNGGRICSCPOSVHSTGECVAGLTQDLRGHSLFEEBSFVLQPSQVSVPFVSVF 1636
Qy 1643 MFTTRQADGVLLQAITRGRSTITTLQREHVMLSV-BGT---GLQASSLRLEPRANDGD 1698
Db 1637 BFRSRAD---MQVF-----ALEFTQRSVHNLEVDGDTLKNMGDBEVELPABEYVSKH 1688
Qy 1699 WHEAQLGASGGFHAHLSFDYGOQRAEGNGLFRLHGLHSNITVGGIPGPAGVARGF 1758
Db 1689 WMNVVILKEAD---SVATISINGIYSAEAKAS--ISDMNLESIFYFIAGPT--GHRSRF 1739
Qy 1759 RGCIGQGRVSPTRPGVNSLDPHSGESINVRQ-----GCSLPDPCD-SNCPANSYCSND 1811
Db 1740 EGCIRNLVLD-----GRSISVKKKKGKTRAGCVPVNCSDVSDICPABS----- 1781
Qy 1812 WDSYSCDPCGYGDNCTNVCDL-NPEHOSVCTRKPSAPHGYSCEPPTYAGPYCETRI 1870
Db 1782 -----DTCLPVSVANVCSSTGCVSNNTA--GYCICIPAGKTGNCCLEA 1825
Qy 1871 -DQCPRGWNG-HPTCPGCNCDVSKGPDPCNKTSGECHKENHYRPPGSPFTCLCDYCP 1928
Db 1826 PKQMCPSGMMGTFPRCRRCSCAQKTDYEACQCKTKGACQCKKSHPTING--CYKCEC-G 1882
Qy 1929 TGSISRVCDPEDGQCPCKPGVITGRQCRC---DNPPAEVTNNGEENVYDSCPRALIEGIM 1985
Db 1883 FGAOSTBES-ADGCKCKNGDAVGRCDRCSPFDQLODSKTLK-CRVSIGKPSIEFSIQ 1940
Qy 1986 WPTTRFGLPAAPCPKSGFTGAVHCDENRGLPPLFNCTISITFSELKFAEELQHNES 2045
Db 1941 WPAQKGSIVYQSCFVBSGLATKCLTETGMSVNMNCRRPREYSIMVNFELFE----- 1996
Qy 2046 GLDSGRSQOALLIRNATQHTAGYFGSDVAVAYQALATRLLAHSTQKGFGLSATQVHPT 2105
Db 1997 -----PSKILTYMANTNTSSIRGRNOQIAAALSLVDEYOSMPMKGAHAKDKMFT 2050
Qy 2106 ENLLR-VGSALLDPAANKRHEMLIOOTBGTAMLQHYEAVASALAKONMHTYLSPTIYT 2164
Db 2051 EKLIESIGRWSEOPADYSTLISK-----LNNYAEVTAETHENV--NFSLPFVAVAN 2100
Qy 2165 PNIVISVVRLDKGNPAGAKLPRYEAALRGEOPDLETTVILPESVRETPPVYRPAQGEA 2224
Db 2101 DHIVASDKLDFGNI---LPKFNMF-----VDLRPTGF-----PRVRAIVAGTT 2141
Qy 2225 QEPPELARRQRHREBELSGEBAVASVITYRTLAGLLPHNYDDKSLKLVPRKPIINTFVVS 2284

```

```

Db 2142 Q-----VYYSIVFY-----PRCRCENPMIA 2162
Qy 2285 ISVHDEBELLPRALDKPYTVQORLLETERTPICYFNMHSILVSGTGMARGCVVR 2344
Db 2163 IVANTSD-----PVIVPEIEEDDMKYPCECRPEK-----SGWTAAAGALLIGU 2208
Qy 2345 NESHVSCQCNHMTSFAYLADVSRRENGEILPLKTLTYVL-GVTLAALLTFEFTLLR1 2403
Db 2209 NLTHACENRIGVFTMFVNDSSSIVRVAQMDNMTSPALAGVALPLCTSLITLSRRS 2268
Qy 2404 LRSNQGIRRNLTALALQVFL--LGINQADLPACTVIALILHFLYLCFSWALLER 2461
Db 2269 LKHSVRIGFILLFPAINIINTLFEVHTALNOA-----YCEVRNAMSFTSSAPFAMFLYG 2324
Qy 2462 LHLVRLTVRDVNTGMEFYMILGKGVPAFTGLAVG-----LDPEYGNPDPCMLSI 2515
Db 2325 LYIYRLAD-----GSSSPSLTSLVGIVPCLLSFTTFPTVTDQCSLS 2369
Qy 2516 YDTLWSFAPVAVAVMSVFLYTLAARASCAARQGFEEKGFPVSGIAPVALLLSAT 2575
Db 2370 HMLFWCILLPLIGLFLLSFYAAASVAVSLHKQYVFAKXNVRAVQHFILITFTG 2429
Qy 2576 WLIALISVNSDTLLFHYLPATNCICQPTPLSYVLSKVRKALKACSRKPS---PD 2631
Db 2430 MTLTGLFANQLPLMEIMEISQSI---YLIALVITLWCVCDDITTKADSDNPSMMLDQ 2486
Qy 2632 PALTKSTLTSVSNCPSPYADG-----LYQYGDASGLSTSRSGSCP----- 2677
Db 2487 KSMVAESTADP-QCASPLSRPHQHREAVPMDSEWPDVNPSTNHT-TSINEPDTIRLL 2544
Qy 2678 -----SYIPFLREBSALNPGQPPGLADPGSL-FLBGQDQDHD 2715
Db 2545 LPQNRDVINILSPDQILBEGVGHVYRNMGSLPRLRSQDEAD 2589

RESULT 3
T00250
MEGF2 protein - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C/Accession: T00250
R/Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A/Title: Identification of high-molecular-weight proteins with multiple BGF-like motifs
A/Reference number: Z14126; MUID:98360089; PMID:9693030
A/Accession: T00250
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1364 <NAK>
A/Cross-references: EMBL:AB011536; NID:93449297; PIDN:BAA32464.1; PID:93449298
A/Experimental source: brain; clone HG104
C/Genetic:
A/Gene: MEGF2
A/Map position: 3p21.2-p24.1
C/Superfamily: unassigned BGF-related proteins; BGF homology; laminin-type BGF-like homo
F/1-28/Domain: BGF homology (fragment) <BGF>
F/32-66/Domain: BGF homology <BGF1>
F/124-169/Domain: laminin-type BGF-like homology <LEG>

Query Match 15.2%; Score 2358; DB 2; Length 1364;
Best Local Similarity 42.3%; Pred. No. 2e-112;
Matches 495; Conservative 180; Mismatches 372; Indels 124; Gaps 25;

Qy 1801 PCPANSYCSNDMSYSCDPCGYGDNCTNVCDLNPCEHOSVTRKPSAPHGTYCECPN 1860
Db 1 PCPNAHDCRDLMTGTECTQPGTIGPCYDACLINFCQNGSSRHLPGAPHGHTCCCVG 60
Qy 1861 YLAPYCETRIDQCPRGWNGHPTCPGCNCDVSKGPDPCNKTSGECHKENHYRPPGSP 1920
Db 61 YFGHHCENRMDQCPRGWNGSPFCGPNCDVHKGFDPNCKNTKGQCKHCFHYRPGSDS 120.
Qy 1921 CLLDCYPTGSLSRVCDPEDGQCPCKPGVITGRQCRCNDNPPAEVTNNGEENVYDSCPRAL 1980

```

```

Db      121 CLEDCDYCVSTSRSCAPHSQCCPCRGALGRQCNCSDFAFVATASGCVLADACPKSL 180
Qy      1981 EAGIMPRTRFGLPAAPCPKSGF-----GTAVHOCENHGMTPMNFCTSTTFELK 2035
Db      181 RSGVWMPQIKETGLATVPCRGALGRGAGAAVALCDEAGCMKBPDLNCTTSIPAFBELSL 240
Qy      2036 PAERLORNESGSDSGSQOLALLRNATOTAGYFGSDVAVAYOLATRLAHSESTORFG 2095
Db      241 LIDGLELNTKALDTMEAKLAQRLREVTGHTDHFSDQDVAVTRILAHILAFSHOQPG 300
Qy      2096 LSAOTVHFTENLIRGSAALLDPAKRMELI-QQTEG---GRAMLIQTEAVASALAO 2151
Db      301 LVTQAHFENLIMAGSALLABETGDLMALQAPGSGAGAGVRIHREKAAATLARN 360
Qy      2152 MRATYSPFTVYNTIVISVRLDKGNF--AGAKLPYEA--LRGEOPDLTTVLTPES 2207
Db      361 MELTYLNPMLVTPNIMLSIDRMEHPSPGARRYRHSNLRGQDAMPHTHVLPSQ 420
Qy      2208 VFERTEPVVAPAGGEAQBELARRQRHPELSQENAVSVIITLGLPHNYDPK 2267
Db      421 SPAPSESEVLTSSSIENSTSSVPPAPPEPEPGISIIILVYRTIGSLPAQOAR 480
Qy      2268 RSLRVRKRPINTPVVISVHDBELLPRALDPVTVQFRLTERTKTCYFNMHSIL 2327
Db      481 RGARLPQNPWNSPVVSAVFRGNFLRGILBSPISLEFLIQTANKSKAICVQMPDPI 540
Qy      2328 VSGTGVMSAGCEVVFVNEHSVSCQCNHMTSPAVLDVSR--NGEILPLKTLTYVALG 2385
Db      541 AEHGVTWADCELVIRNGSHACRCRGTTFEVLNDASREELBGLLAVTHVVA 600
Qy      2386 VTLAALLTFPFLTLRIKSNQHGIRRNITAAQLVGLGINQADLPACTYAIL 2445
Db      601 VSAVALVTPRAIILSRKSNVGRGHANVAALGVALFLGIRHRTNQVCTAVAIL 660
Qy      2446 LHELVICTSMALBELHLYRALTEYADVNTGPMRPTMNGVAPRTGLANGDPEGY 2505
Db      661 LHFVFTSTAMLVVQGLHLYRMQVEPNVDGRMRFTHALGWGVPVVLGLAAGDPEGY 720
Qy      2506 GNDPFCMLSTYDTLINSFAPVAVAVSMVFLYTLAARASCA--AORGEFKKPGVGLD 2564
Db      721 GNDPFCMISHEPLINSFAPVAVVIMNGTMTLAAKRTSCIGQEA--KTSALTLS 778
Qy      2565 SPVAILLSATWLLALLSVNSDTLLPHYLTPATNCIOGPFIPLSYVLSKRVKALKAC 2624
Db      779 SFLLLLVSSMYFLGLAVNHSILAFHYLAHGCGLAVLILFVLNADAAAMWPA 838
Qy      2625 -SRKEPDPALTKSTLTSYNCPSPADGRLYQ--PYGDSAGSLASTSGSKQP----- 2677
Db      839 LGRKAPEERAPRGLGPGAYNTALFEESGLRITLGASTVSVSARSGRTODDSOR 898
Qy      2678 --SYI--PFLIREESA-----LNPQGPFGG-----DPSGLFLEGDDQHDPTDS 2720
Db      899 GRBYLDNVLVIRGSAADHTDHLQAHAGPTLDVAMFRHDA-----DSBS 946
Qy      2721 DSDLSTEDQSGSYASTHSSDEEBEEREAAPGEQCMDBILPGARLPLHSTPX 2780
Db      947 DSDLSTEEERSLSPSESEDNRTGRFORPCRAA-----SERLLTH--PCD 994
Qy      2781 GCGGPGKAPRG-----DGT-----TAKSSGNGABERLREKDALSR 2821
Db      995 VPDNDLLSVPALGECBAAPCALQTWGSERLGLDTSKAAANNQDPPL--TSGD---E 1049
Qy      2822 GSLGPPGSAQPHKGLKKKCLPTISEKSLRLP-----LEOCTSGSRSSA- 2870
Db      1050 TSL-----GAGQRKGLIKR-----LQYPLVPTRGAPELSWCRAATLGHRAV 1094
Qy      2871 -----SESGRGPPEPPRQSLQEQIN 2893
Db      1095 PAASYGRIVAGGTGSLGPAASYSSREQLD 1125

```

---

```

IUFTM
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #ext_change 16-Feb-1997
C:Accession: A41087; B41087
C:Name: P.A.; Weber, U.; Onofrechuk, P.; Blessmann, H.; Bryant, P.J.; Goodman, C.S.
Cell 67, 853-868, 1991
A:Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
A:Reference number: A41087; PMID:92069752; PMID:1959133
A:Accession: A41087
A:Molecule type: mRNA
A:Residues: 143-485;1279-5147 <MAH>
A:Cross-references: GB:M80537
A:Accession: B41087
A:Molecule type: DNA
A:Residues: 1-142;487-1278 <MA2>
A:Cross-references: GB:M80537
A:Note: 1229-Gly and 1233-Ser were also found
C:Genetics:
A:Gene: fat
A:Cross-references: FlyBase:FBgn001075
C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-5147/Product: cadherin-related tumor suppressor #status predicted <EXT>
F:36-4583/Domain: extracellular #status predicted <EXT>
F:51-156/Domain: cadherin repeat homology <CR1>
F:159-270/Domain: cadherin repeat homology <CR2>
F:271-382/Domain: cadherin repeat homology <CR3>
F:390-494/Domain: cadherin repeat homology <CR4>
F:497-599/Domain: cadherin repeat homology <CR5>
F:602-708/Domain: cadherin repeat homology <CR6>
F:718-822/Domain: cadherin repeat homology <CR7>
F:831-942/Domain: cadherin repeat homology <CR8>
F:948-1049/Domain: cadherin repeat homology <CR9>
F:1052-1153/Domain: cadherin repeat homology <C10>
F:1156-1278/Domain: cadherin repeat homology <C11>
F:1281-1384/Domain: cadherin repeat homology <C12>
F:1387-1489/Domain: cadherin repeat homology <C13>
F:1492-1601/Domain: cadherin repeat homology <C14>
F:1607-1713/Domain: cadherin repeat homology <C15>
F:1717-1823/Domain: cadherin repeat homology <C16>
F:1826-1922/Domain: cadherin repeat homology <C17>
F:1925-2027/Domain: cadherin repeat homology <C18>
F:2028-2167/Domain: cadherin repeat homology <C19>
F:2169-2278/Domain: cadherin repeat homology <C20>
F:2281-2384/Domain: cadherin repeat homology <C21>
F:2387-2491/Domain: cadherin repeat homology <C22>
F:2494-2596/Domain: cadherin repeat homology <C23>
F:2599-2703/Domain: cadherin repeat homology <C24>
F:2707-2810/Domain: cadherin repeat homology <C25>
F:2813-2913/Domain: cadherin repeat homology <C26>
F:2915-3013/Domain: cadherin repeat homology <C27>
F:3014-3124/Domain: cadherin repeat homology <C28>
F:3127-3229/Domain: cadherin repeat homology <C29>
F:3232-3334/Domain: cadherin repeat homology <C30>
F:3337-3439/Domain: cadherin repeat homology <C31>
F:3442-3545/Domain: cadherin repeat homology <C32>
F:3548-3651/Domain: cadherin repeat homology <C33>
F:3654-3756/Domain: cadherin repeat homology <C34>
F:3954-4010/Domain: EGF homology <EG1>
F:4017-4048/Domain: EGF homology <EG2>
F:4056-4089/Domain: EGF homology <EG3>
F:4096-4127/Domain: EGF homology <EG4>
F:4584-4609/Domain: transmembrane #status predicted <TM>
F:4610-5147/Domain: intracellular #status predicted <INT>

```

---

```

Query Match      10.0%; Score 1549.5; DB 1; Length 5147;
Best Local Similarity 22.8%; Pred. No. 3.3e-70;
Matches 553; Conservative 293; Mismatches 674; Indels 903; Gaps 71;

```

---

```

Qy      161 RSPESLGRGRKRVNTLPQFPSPYQATVPENQAGTPVASLRALDDPBGAGRLTYTM 220

```

Db 2258 QSTESQLVRLIDENDNSPVFDPKOYSASVAENASIGAMVLQVSAITVDGANGRIRYSI 2317  
 QY 2221 DALFDRSNQFSLDPYTGAVTTABE----- 246  
 Db 2318 -VIGD--QNHDFISSEDTGVVAKUNLNERLSRLTTRAEPCALBNPAGDTAELTINI 2374  
 QY 247 ----- 246  
 Db 2375 LDINDNRPTFLDSPYLARVMENTVPPNGGYVLTVNAVDTADPTPLNSQVRYFLKESGDLJE 2434  
 QY 247 -----LDRETSVFRVYTAODHGMPPRSALATTLVLTVDNDHPVPEEQ 292  
 Db 2435 RINASGDIALLKPLREBOSEKYLTLVAMDTGSPPLTGIVRVEQDINDNDPVFEIQ 2494  
 QY 2293 EYKESLRENLEVEYEVLTATVATDGDAPPNANILYRLLEBSGSGSPSEVFEIDPRSGLYRTR 352  
 Db 2495 SYHATYRENUPLSGTHVLTTPATDDBGLAKAPNL-----GERMERFHIIDSTGEISIA 2550  
 QY 353 GPVDRSEVESYOUTVEASDQGRDPGPRSTTAAYFLSYEDNDNAPQFSEKRYVVOYREBY 412  
 Db 2551 TTLDRETSVYHLTLMAQDSITE-PRASSVNLTISVDVNDNIPKDSSTTVNAVAPERI 2609  
 QY 413 TPBAPLAVTASRDGKSNAVHYHYSIMSGNARQPTLDAQT----- 453  
 Db 2610 SKGEFVGARALDLDDGENAVVHYTI---SGRDQHYFDINTKTVGVSTKLELTKTKSTD 2666  
 QY 454 -----GALD----- 457  
 Db 2667 DLTTTVISAMDQEGSLSKAELTVLRPELPPTPAVANSHTMSEVBRGKMTTKY 2726  
 QY 458 -VVSP-----LDYETTKYTLVRAQDGG 480  
 Db 2727 SATSPKGLVGIKIRYALAGIMGSLRVDPMGSLSVQDGLDYELTHLYEIMWBAADG 2786  
 QY 481 RPLSNVSGLVYTVQVLDINDNAPITVSTPQATVLESGVLYVHQALDADAGNARL 540  
 Db 2787 TPLRSTLT-LITNAVTDANANAPMEQLITAYALEBSPQILAVKASDRSGNGNV 2845  
 QY 541 EYRLAGVGHDFPFTIN-NGTGMISVAALDREVEYDFSGVEARDHGTPLATASASVY 599  
 Db 2846 IYHLQ---NDFDGFELTESGEIYTRRLDRBELGDVAFVEAVDQVPHLTGTAIVLH 2902  
 QY 600 VLDVNNPFTTQPEYTVRLMEDNAGTSVYVTSANDRD--AHSVITYQITSGNTRNRS 657  
 Db 2903 LILKNDNPPKPTR-LFSLANTENAKISFVIRVTSPLDGLGANANASYSF-SEMPGKFR 2960  
 QY 658 ITSOSGGLVSLALPLDYKLERQVYLAVTASDGRQDTAQIVNVVDANHRVPOSHY 717  
 Db 2961 IEPQSG--ITVAGHLDRQODEYILKVASDAMRAETPTITTIQDQNNAPFEHSFY 3018  
 QY 718 TVANVEDRPAGTVVVLISATDEB-TGENARITYFMEBDSIPQFRLADDTG----- 765  
 Db 3019 SPSPELQSIALVGOIATDRDKQGNVSIYSIQPSMPSIDPATGVSPEKAVRFK 3078  
 QY 766 -----AVTQA----- 771  
 Db 3079 HSQVPSBENMYALTVALTNGKPEPLXSECLVININIVDAHNPKEQOAEYLAPLQDAV 3138  
 QY 772 -----ELDY-----DQVSTTLA 784  
 Db 3139 RQGRIVRHANDKODLTENEDYSLMTFNLSIFSVERHDMITLVKPIOVPPNTRJELV 3198  
 QY 785 ITARDNGIPKSDPTTYLEILVNDVNDNAPQFLRDSYSGSYVEDPPTSVLOISATDRS 844  
 Db 3199 VRATDRGVPPQSDETRVIIVTGENMDTPRPSVNSYVIVPENPVGSTLITVGTATDDT 3258  
 QY 845 GLANRVTYTOQGDGDGDPVVESTSGIVKTLRLDBENVAQVYLAAYADKMPAPATP 904  
 Db 3259 GPNMGMLAYSISGNERO-DFSVDERTGVIYQOQLDLDLIOEYHLNLTVDLGIHPLSSV 3317  
 QY 905 MEVTVTVLDVNDNPPVEODEFVFEVENSPIGLAVARVATTDDEGTNA----- 954  
 Db 3318 AMLTITLTDVNDNPPVNRHKEHYCIPEKNFVGIVTQAHATDKDSKRNAILHAPLPSG 3377

QY 955 ----- 954  
 Db 3378 PDRHEPIMQNSGTISSAVSPDYERRIYTLQIKANPDSSMESTANLYVHVLGVNPEFP 3437  
 QY 955 ----- 954  
 Db 3438 QFLQVPHFVDSAVGTVRGAVQATDKSGEDGRVYLLVGSSNDKFRIDTWTGLY 3497  
 QY 955 ----- 954  
 Db 3498 VARHLDRETQNEVLTVMANKYSGIRGNDTDEAQVITISIQDNDEPEIKHYTSTISBA 3557  
 QY 955 -----QIMQYVEGNIPEVFOIDIPSGELTALVDIDYEIDREPY 992  
 Db 3558 AVGVKTTVKAIDDVRTQNNQPSYSIINGLQSPKIDVGTGHSITSRIDRETSY 3617  
 QY 993 VLVIOATSA--PLVSRAVHVRLDRDNDPVL--GNFELFNNTYNNRSSSPGCAL 1046  
 Db 3618 NIVIGALIDTGLPQTSATVHIELEDVNDNGPTPEG-----LNGYISENEPA--GTSI 3670  
 QY 1047 GRVPAHDPI--SDLSYSPERGNELSVILNASTG----- 1080  
 Db 3671 MTLMASDPDLPRNGGPFYTLQIGKHKSWLSDVNSGVVRSSTTSFDRMTPLVLAIVE 3730  
 QY 1081 ----- 1080  
 Db 3731 DSGKPKQSQHLITTVLDQNDNPSTSLHIAVSLFNGDLPSNVKLVADVRNDIDYGD 3790  
 QY 1081 -----EIKSRALD-----NRPLEAIMSVLSDGVH-SVTAOCALRVTTI 1120  
 Db 3791 YRCRLQKNPAQOLQALIPRACDLITTSHTTYIAVFSYTGDKGKGDVSVSVAFOSF 3850  
 QY 1121 TDEMTHSTILRLBMSPERFLSPILGLPIQVAAVTLTPPHVYVFNVRQBDTAPGHI 1180  
 Db 3851 NNKETLANSVIVNRNNTANHFLANHTRPLIEIKRMSN-DEEVILYSL--LEGGSGNS 3906  
 QY 1181 LNVSL-----SVGPPGPGGPPPLPSBLOERLYNLSLTLTASQRYLPEDDNI 1231  
 Db 3907 TNLQILMAVRLAKTSYQD-----KYLIERLEKKSASFELLOKEVIVGYERC 3954  
 QY 1232 CIREPCENRCVSVLRFPSSAPFIASSS--VLRPHIVPGGLRCRCPGPTGD----- 1283  
 Db 3955 SBDVCENGVCSATMRLIDAHSPVODSPALVLSGP-RVHVDYSQCTSGSGSEGCRR 4013  
 QY 1284 -----YCEB--VDLCYSRPGPHGR-C-BSRBS 1308  
 Db 4014 QDPCLPNCHSOYQCRRLASDPOCMCPANRDGKCHKERSDYCYSKPCANGSCQRSPEG 4073  
 QY 1309 -GYTCLRDGYTGERCEVSARSGRCTPGVCXNGGTCVNLVGGFKDCPSGDFEKEYQV 1367  
 Db 4074 SSYFCLCRGFRGNQCE--SVBDCRPNPCLHGLCVS-LKGYKNCCTPRGYGR-HCER 4129  
 QY 1368 TTRSPANSFITPRGLRQRFHTLALSPATKRDGLILYNGFNE-KIDPALAVIYQO 1425  
 Db 4130 FSYGPOPLSYMTFPALDVTTN-DISIVFATYTPNSLILNYCMQSGRDFLAIBLVHR 4188  
 QY 1426 VOLTFASGSTTVSPFPVG-GVSDQOMHTVOLKTYNKKLLAQTGLPQSPBOKVAVTV 1484  
 Db 4189 AY--PSSGARTALISVIVAGRLADOGMKV-----TATRNG-----RWSLSV 4230  
 QY 1485 DGDGTVALRFGSVIGANSYCAAQ-----GTQSGSKSLDTGPLLLGVY---PDLPSF 1535  
 Db 4231 AKCADSGDVCTECLPDDSCYADEVAPVGTILFNKQ-----PLMIGLSADPILERP 4284  
 QY 1536 PYRMRFVGCMMNLQVDSRHIMADPIANGVYPGPAKQVCD-----SNTCHN-GCTV 1590  
 Db 4285 QVHSDDLVGLCHSVHIGGALNLSPLQOKGILAG--NRQACOPALAEKRGCFAGQCI 4342  
 QY 1591 NQMDAFSCEC-----PLGFGKSCAQ-----EMANPOHFLG--SSIVAM-- 1627  
 Db 4343 DRKSSSLCQGGHLSQSDCSLSLEPITLGBGAFVERIBSEIYRBMQLUNLNKSKAMLD 4402

QY 1628 -----HGLSLPISQPMWLSIMFRTQADGVLLQATIRGRSTITLQREGRHW 1674  
DB 4403 NOOMERRAVNSFSTASQIYEAKPMISMLFRFKDKQOQILIAATNOMFT-SLSLRGRLV 4461  
QY 1675 -----LSVEGTGLQASLSLLEFGRANDGDMHHAQALGASGGPGRHAILSFVQOORA-- 1726  
DB 4462 YSKOHLTIMTVOERSTL-----NDGKMHVSL-----FSEKSLR 4498  
QY 1727 ---EG-NLGRHLH-GH-----LSNITVGIIPGAGVARGRCLOGVRVSDTPEGV 1774  
DB 4499 LAYDGOVQDELDIAGVHDPDLDEYLTILNVG-----BAFVGLANVTYNNELQPL 4549  
QY 1775 N---SLDPS---HGSEINVEQC 1791  
DB 4550 NGSSTIFPEYRYHCK---IESGC 4569

RESULT 5  
T30213  
G-cadherin - sea urchin (lytechinus variegatus)  
C:Species: Lytechinus variegatus (variegated urchin)  
C:date: 22-Oct-1999 #sequence\_rev199 #text\_change 22-Oct-1999  
C:Accession: T30213  
R:Miller, J.R.; McClay, D.R.  
Dev. Biol. 192, 323-339, 1997  
A:Title: Characterization of the role of cadherin in regulating cell adhesion during sea  
A:Reference number: Z20780; MID:98104238; PMID:9441671  
A:Accession: T30213  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2809 <M11>  
A:Cross-references: EMBL:U34823; MID:g2982186; PID:g2982187; PIDN:AAC06341.1

Query Match 9.7%; Score 1508.5; DB 2; Length 2809;  
Best Local Similarity 29.0%; Pred. No. 1.7e-68;  
Matches 538; Conservative 252; Mismatches 757; Indels 311; Gaps 87;  
QY 147 SCKLAQBPGLRAGE-----RSPESLGGRRKNNTAPQOPPSYQATVPENOPAGTP 199  
DB 938 SCTIVASDSIRSRVDVDFVRSPPDT-----TYQVPTSPNTPYQVTINMENTTIG-D 988  
QY 200 VASLRAIDPDEGAGLFTYMDALPDSR-----SNQFSL-----DPYGAVTAAELDRPT 251  
DB 989 VILLVSV-----GAQEVGPDIVKSRSTNSDGTFTLTDDNLQSDLTIVKLDPEA 1042  
QY 252 KSTHFRVTAQDHGMRRSALAT-LTILVTDTNDHPVEQOQYKSLRENLVGYE--V 308  
DB 1043 VQEVLELTVAQYVGVQBFIAQPTVLTIEDFNDVAPKPAVTPGNIENQPVNSDPI 1102  
QY 309 LTVKATDGAAPPNANILYLLSGSGSPSEVBEIDPRSGVITRGFVDRREVE--SYQLT 366  
DB 1103 IYLOATIDGVTPE---YSQISYSPNABSDSIDETGEIFALKVDFEALIEDSIALT 1158  
QY 367 VEASPOGRDPRGRSTTAATFASVEDNDNAPQFSEKRYVVOYREDVTPAPVLTATASR 426  
DB 1159 VASD-----GNODLATTFTITVDENDEPEEN-GTSPFVLEEDGKMGYDGTATADD 1212  
QY 427 DKGSNAVHYYSIMSGNARQFYLDQGTGALDVVSPLDYETKRYTLKRYAODGGRPLSN 486  
DB 1213 D--ISEVLEYPIISGNEGGAFVDAEGTIRKAGVLDYEAFTSYELQSVVNDG-----KN 1265  
QY 487 V-SGLVTVQVLDINDNAPFVSTPPOATYLE--SVPLGLVLAHQAI--DADAGNARLE 541  
DB 1266 VATTVTVTINLVNNDVAPQPDQASASASYIEBDNSLPRILLSVATDQDADAVDA--V 1324  
QY 542 YRLAGVGHDPFTINNCTMISVAELDEEVDYFSFGVEARDHGPALPTASASVYVL 601  
DB 1325 YELVGTGACTITLISQITNTITLQALDREKIFPTINLAVAATDNGNLISYVDVTIYE 1384  
QY 602 DVNDNNPFTQDEYTVRLNEDAAVGTSVTVASVDRDAHVTITQITSGNTNRSPTSQ 661  
DB 1385 DINDNAPVPPDQEVGSVEENRPPNTPVVAVAABDPTADDLMTSPPTSP--DENINSQ 1442

QY 662 SGGGLVSLALPDYTL-ERQYVLAATASDGTRODPAQIVANVTANTHNPVFOSSH-Y-TV 719  
DB 1443 T--GGITAKQDREPTPESEIEIVQATDGVNTASTATVVISIDVDNMPSESEVDYPPDA 1500  
QY 720 NNNBRPAGTTVVLISANDEDEGENARITYFEBSDIPO--FIIDDT-----GAVTQAL 773  
DB 1501 SVLETEPIGTTITVQALIDPDVDFRDAVEFSINSQDPBELFPIVADAAITQGLIQVNNKL 1560  
QY 774 DYEDQVS--YTLATPARDNGIPEOKS---DTYLAELVNDVNDNAPQFLADSQGSVEYEDV 828  
DB 1561 DETLATNREFLTVAVTDSQGTDSGRPETAVKLIYENANDLAAVFQDQYDGAVERSEA 1620  
QY 829 PPTTSV--LQISATRD--SGLNGRVFTYQGGDGDGDFIVSTSGIVRTLRRLREN 883  
DB 1621 TVNSQVGTFTFSATGDEPRGATLN---YIIDPTDPGQFAINNOQLI-VASPLDRRT 1675  
QY 884 VAOYVLRAYADKGMPPARTMEVTVTVLDVNDNPPVPEODEPDPVPEENSTI--GLAVA 941  
DB 1676 VASIELKAYADVNGPSPSGTATVAVTVTDVNDTPHFAQD--YNPSVEE-GRLEANVEV 1733  
QY 942 RVLTADPDE-GTNAQIMQIVEGNIPEVQDLIF-----SGEL---TALVDLYE 987  
DB 1734 SVEAVDDDDPPSGPFLNVA-----PQNDWTFPDIDGLGTSGSIRVSTYGLIDRE 1789  
QY 988 DRP--EYLVITQATAP--LVSRATVHRLDDRNDPVLGNFELFNNTYNNRSSSPG 1043  
DB 1790 THPYEDIVFLAEVGTPEALNGTQTLTIMISDVNDNPHVATIKOILVYSBGN---IPT 1845  
QY 1044 GAIGVPAHDDDISLTYSPFERGNEI--SYVLNASTHEKLSRLDNNRPLEAI--MEVL 1101  
DB 1846 TEVGAVGVEDPDILDKTY--EAVGELPDFOLDSDTDTMA-----EGTPAGYEMDIR 1899  
QY 1102 VSDG--VHSVTAQCALRTYITIDEMLTSTITRLSDMSPERFLSP-----LGLIF 1149  
DB 1900 VSDQSTYSVSVTVIVYKQIPREAVFSSGSIRSGTAAELITPDSGVSNDLKTIL 1959  
QY 1150 IQAVATLATPPDHVNVN--ORDTDAP--GSHLVNLSVGGPPGGGPPPLPSDL 1205  
DB 1960 ANAVCAQLAN-FDIFSVLVNVMGMRVDIRYAAH-----GSPYPAQL 2002  
QY 1206 QERLYANSLTLTAISAQ-----VLPPDNICLR--PCRNVMCVSLARFSSAP 1254  
DB 2003 D-----LALSVSNEIEELGISTAQIPVD--LCVENVCBS--SCTNVLVVDPIPT 2049  
QY 1255 FIASSVLEFPIHPYGLRCRPPGFTGYCETVEDLYCSRPGPHGRCSRREGYTCLC 1314  
DB 2050 VVDSGTALVAITSV--LEAQ-----ICGANTVAP----- 2078  
QY 1315 RDGTTGHECVSASAGRCTPEVCXNGGTCVNLLVGGFKDCPSSGDFEKPVCQVTTSPFA 1374  
DB 2079 -----GNCDSDPCINGGTCEBDVHGTYRCTPY-LFDGPNCOQTKRSP-E 2121  
QY 1375 HSFITPRLGRPHFTLALSPATKERDGLLYNRP-----NEKNDPALFVIOEQVLT 1429  
DB 2122 NGYASFSTLRCERTSLSIEFTYVSSGTLNBPFIPTPGDDPIDMLLELIGKALYT 2181  
QY 1430 FSAGSETT-----VSFPVGVGVSDQGMHTVOLKYNNKPLAQGTGLPGSBEOKVAVVT 1483  
DB 2182 INLSTGSTDNVLBAVPTDQTQNDNEMHRIIDV-YNN-----GRFVEMT 2225  
QY 1484 VDCGDTGV--ALRFGSVLYNYSQCAQGTQGSK--KSLDITGLULLGV-PDLPESEFPV 1537  
DB 2226 VDR-C-MGFPAETSSSSTLDDSSCRVNGTTPGKNNLFLNHTPLFLGMSADVDVVP- 2283  
QY 1538 RMQPFVGMRLQVDSRHIMADPFIANGVTGPGPAKKNVC--DSN-----TCNNGGCVNQM 1593  
DB 2284 --SGFDGCIKULVSDGFLYDGTGTSKSEAGCPRIQDQCTDNGMPVCNN-GTCEADI 2340  
QY 1594 DAFSCEPPLGFGKSCAQEMANPQHLGSSLVAMHGLSLPI---SQPMYLSIMFRTQAD 1650  
DB 2341 DSFICICPPGNGLTQVLEL--TFYDAIBSYITYELDDSLYDARSNNYQIMVTRQEN 2399  
QY 1651 GVLQALITRGS--TITQLRBEHVMLSVBGTGLQASLSLLEPGRANDGMH-----H 1701

```

Db      2400 G-LIMSSISANTYEMIMVQGLKADWH-LGDKPVSVMNFSINDGMAHINFDKXD 2457
Qy      1702 AQAALASGGPGRHAILSPDVGQGRABENGLPRLHGLT-SNITVGGIPGAGVANGPFG 1760
Db      2458 SVTITIKIDGGGKVEKIO-----NRESQYSGLDVDENSIVIGAFVDNVTITDDPMG 2507
Qy      1761 CLQGVAVSDT---PEGVNSLDPHGESINVEQCSLPDPCSNPCPAN-SYCSNMDSYS 1816
Db      2508 CMMDPRIINHHFLMGEBNDNAVA-TKSAGTBGP-SDVCDSDPCPSGILVCTDYRPFIE 2565
Qy      1817 GSCDPEGY-----GDCTNVCIL-NPCHQSVCTRRKSPAPHGTCPCPNYIGPYCE 1867
Db      2566 CLCPGEGEVEEDPDCTMAIIDCVPMFCANGTCV--GDPTGTCDCPSGYGDRCE 2621

RESULT 6
T00021
DN-cadherin - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 17-Nov-2000
C/Accession: T00021
R/Iwai, Y.; Ueno, T.; Hirano, S.; Steward, R.; Takeichi, M.; Uemura, T.
Neuron 19, 77-89, 1997
A/Title: Axon patterning requires DN-cadherin, a novel neuronal adhesion receptor, in th
A/Reference number: Z14058; MUID:97388431; PMID:9247265
A/Accession: T00021
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-3097 <IWA>
A/Cross-references: EMBL:AB002397; NID:g2381491; PIDN:BAA22151.1; PID:g2381492
A/Experimental source: whole embryo and adult head
C/Genetics:
A/Cross-references: FlyBase:FBgn0015609
A/Map position: 36D
A/Note: 11(2)36DA
C/Superfamily: Unassigned EGF-related proteins; EGF homology
F/2346-2377/Domain: EGF homology <EGF>
F/2869-2902/Domain: EGF homology <EGF>

Query Match      9.1%; Score 1421.5; DB 2; Length 3097;
Best Local Similarity 24.7%; Pred. No. 5.5e-64;
Matches 514; Conservative 286; Mismatches 747; Indels 533; Gaps 72;

Qy      179  POFQPPS-YQATVPENQAGTPVPSLRALIDPDEGAGRLTYMDALFDSRNOFFSLDPV 237
Db      966  PVFQDSTYTPKVEGAPNSPVIKVATDEKGVNQVYSIQGNQGTGK-FTVDEE 1024
Qy      238  TGAVTTAEBLDRFTKSTHVRVT--AODHGMPPRSALATITLVTDTNDHPVPEQGEYK 235
Db      1025  TGRVSTKGVFDRREGDDKPFVSVTKATDQGPDLSEGVCSFTVBITVDNDPPLFDRQKYV 1084
Qy      296  ESLRENLEVGVEVLTARATGDAPAPNANILYRLLEGSGSPSEVFEIDPRSGVIRTPGPV 355
Db      1085  ENVAQDSIGTINILKRVASDEBDADNNGAIVISLAPNPDLBEFELQAESGMITLKKEL 1144
Qy      356  DREVESYQUTVEASDQGRDPGRSTAAVFLSYEDNDNAPQSEKRY-VVOVREDVTP 414
Db      1145  DR---EYTKLEANAQDGYR--PLSRVTEQIDVVDANPNPVMWDHVVGPITVYKEMPV 1199
Qy      415  GAPTATVTAAGDRDGSNAVHYISMSG-----NARGQFYL-----DAQTGA-LDVYSPLD 463
Db      1200  GGRKVASIKASSGREG-NPTVYRLMPGSTAQTNNKFFYLLQQRDNEDTVAADIKVNHPLD 1258
Qy      464  YETTKETLRLVRAQDGRPLSNVSGLVTVQVLDINDNAPLIVSTPQATVLSVPGLYL 523
Db      1259  YESIKETNLTRVANNAGQLAS-BATVYIMLEBVNDIEPLFTRE-QEYVLSEEPITGK 1316
Qy      524  VLVHQAIDADAG-----DNARLEKRLACVGHDPPTTINNGTGMISVAELDREE 572
Db      1317  VTQNALIDKQGFENNQYVYIYVDSARNE-----GKEF-FEILQSGEIFTKVVFDRK 1369
Qy      573  VDFPSFGEVARDHGTPLATASAS-----VSATVLDVANDNNTFTQPEYTVRLNDDA 624

```

```

Db      1370  KCAVALLEVEARD-GAPSARPNNSNGNSVTKFRIIGLAKNDNPFPYDLSYAEADENED 1428
Qy      625  VGTSVVTAANDRADHSVITTOITSGNTRNRSINSQSGGLVSLALPLDYLERQYVLA 684
Db      1429  IGHVLTATYAKDHSSIRHETISGNTIGAPAVNMGTGA--TVAGALDIETRRRIELR 1486
Qy      685  VVASGTDQDTAQIVANVTANDTHRPVFSHYTVANVE-----D 724
Db      1487  LAASNLKENTYTTVLIHKVDVNDNPVERPFRYQITEBDRNLPKGYLQVATDGDKD 1546
Qy      725  RP----- 726
Db      1547  RQNTIVYELTQGIIDPNPNANSKPDINRTGEIFVLKELDRDQPNGRPQMRFEVADQEG 1606
Qy      727  -----AGTVVLISATDED-----TGENA 745
Db      1607  GEGLVGYADVQVNLKINDNAPIFPQGYVFGAVTENGTAAGVMTMTAVDITDDPNEGNA 1666
Qy      746  RITYEMEDI-----PQFRIDADTGAVTQA----- 771
Db      1667  RLVTYSIEKNVIEETGSPILFEIBPTGVILKTAVCCLDBERFPDYSIQVAMDGGLKGTG 1726
Qy      772  ----- 771
Db      1727  TASIRVKOINDMPQPTKDEWETFEVDETDGALPEMPLITVTHDEDTNKFQYKVIDNS 1786
Qy      772  -----BLDYEDQV-----SYTLATFARDNGIPQKSDT-----TY 800
Db      1787  GYGADKFTWRNRNDGTGSLIKIVQPLDYEDQLOSNGRFRIRIQVNDGENDNKKHYAVSW 1846
Qy      801  LELVNDVNDNAPQFLRDSYQGSYVDEVPPTSVLQISATDRDSGLNGRVFTYFGGDDG 860
Db      1847  VVVKLRDINDNKHFERANVAVSEFEDTKVGTLEKFKATLDPDQGGKSVSYISRSBR 1906
Qy      861  DDDFVSESTSGVRLRLDRDENVAQVYLRANAVDKMPRATPMENVYTVLDVNDNPPV 920
Db      1907  QROFAL-NONGSVTLQSLDRVVRHGVKILALIDGSPKATATLTLYIVQDINDNAK 1965
Qy      921  PQDEEDVVEENSPIGLAVATATDPDEGNAQIMYQIVGNIPVEFQDIFSGE--- 977
Db      1966  FLKDRPVLPEHVPP--KRVSEILATDDDRKS-----NGPFGRLDPSADDIR 2015
Qy      978  -----LVALVDLYEDRPYV--LYIQATSAP-LVSRATVHRLID 1015
Db      2016  ASFKYEQDQKANGDGMVAVISSLSRDFDEQKRWIMPIYIKDGPAMTGSTLTLYIIGD 2075
Qy      1016  RNDNPPVIGNFELFNNTVYTRNSSSPGGAIGRVPAHD--DISSLTY-----SPE 1065
Db      2076  VNDNKKQPSKQIDFYVNY---QGQSPDTPICKRVYVYLDMDLDPDKFTYEMAMHPRK 2131
Qy      1066  RGNELSLVILNASTGE-----LKLRLADNRRPLEAIVSLVSDGVH-SVTAQCALRYT 1118
Db      2132  LDEDSGMTVRAGTREGRHIAFKYDRKHGTQDIPANTVTVRELPHEAVNNSGVRLS 2191
Qy      1119  ITDEMLTHSTLRLEDSPPERFLSPLGLFQVAATLATPDDVAVVNVYQDIDABEG 1178
Db      2192  GISDEDFIRVWNYKQMSRSK-----MDRFQDKLADLINTREAVDIFSVLKRKP-- 2244
Qy      1179  HILNYSLSV-GQP---PGGQGPPLPSEDLQRLYLNRSLTALASQGLVLPFDNICKR 1234
Db      2245  PLTDVAFSAGSPYKFPVRLANGIYLMHREIKVDGIN--ITMGIDBCL-YENQMC-E 2299
Qy      1235  EPCENYRCVSVLRPDSAPFIASSSVLFRPIHPVGLGRCPGFTGDCETEVDLCYS 1294
Db      2300  GSCNLSLEI-----SPLPYVNNAN-----KTALVGRVDTIAD----- 2332
Qy      1295  RQGHGRGRSRBGGTYCLCRGYTGENHCEVARSRGCTPGVCKNGCTVNLIVGFKCD 1354
Db      2333  -----C-----TCGARNFTKPESC-----RTTP--CHNGKRCVDRFGB-HCS 2367
Qy      1355  CSPGDEKPYCQVTRSPFASFTFRGLRQRFHTLALSPATKERDGLLYNGHFNEKH 1414

```



```

Db      2368  CPVQ-YTERCCQQTTRSFNGMAMPYLEMCDESHLSLEFTRKPDGLIYNGEIPPE 2426
Qy      1415  -----DFVALEVIOEQVQLTFSGESTYTVSPFPGVSDGQMTVOLKYNRPDLGQ 1467
Db      2427  RDETLISDPIALELBERGVRLLIDPGSSGLBELRVKTKLIDGEMHRIDLPMDTESIR-- 2484
Qy      1468  TGLPGPSBQKAVAVTVTDCC-----DTGVALARGVSLVNGSCAAQGTGSGKKSIDL 1519
Db      2485  -----NWVDFCKSAELAEEMEDGTPPEFDW-----SCQARQIIPFNEYLANV 2526
Qy      1520  TGPLLGGVPLDLPSPFVPM-----ROFVGCGRNLOVDSRHIDMADFLANNGTYPG 1570
Db      2527  NAPIVQGLI--YRQPDOSLYRMTHTPTAKGDFGCTRLVHNSKLYDLAHPGLSRNSVAG 2584
Qy      1571  CPKAVNC-----DSNYCHNGTCVANOQMDAFSCCEPLFGGKSCAQEMANPQHFLSSLYA 1626
Db      2585  CQTEVCAQGTETTRACMCHGNCVGLSBRACHCRGWTGPAC-NIPITPTFKAQSYVK 2643
Qy      1627  WHGLSLPISQPWYLSLMTFTROADGVLLQATIR-GRSTITTLQLRGAV--MLSVETGLQ 1683
Db      2644  YALSEEPFRFTQVOLFRTREERYGELFRVSDQHNREYGLIEIKDGHILFRYNLNSLRT 2703
Qy      1684  ASSTLEPGRANDGMHHAQLALGASGGRHAISFDVGGQRAEGN-----LGR 1733
Db      2704  EKDLMLNLIYVNDGQMHVVKVRYSAA-----ATLELDGEGRRNTEPEVGHQMLVDK 2759
Qy      1734  LHGLLSNITVGGIPGAGVA-----RGCLQGVAVS-----DTEPGVNSL 1777
Db      2760  QEGVY-----AGKAEYGVRTFRVYADYQKCLDILYEGKHLPLPAMNG- 2806
Qy      1778  DSHGSEI---NVEQGCSLPDCDSNCPDANSYCSNDSDYSQSC-----DPGYGNC 1828
Db      2807  -TQMCAQTMARNLKRCGSCNKCNSNVTCPDPCEVDLNMVVECTGCBGRIMSPDKGMD 2865
Qy      1829  TNYCDLNPCHQSVCTR-KPSAPHYTCCECPNVIQPCPE 1867
Db      2866  RNECLDMFCMNGATCINLEPRL-RYRCICPDGFGENCE 2903

RESULT 7
100252
MEGF1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00252
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomic 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:963030
A:Accession: T00252
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4351 <NAK>
A:Cross-references: EMBL:AB011527; NID:G3449285; PIND:BA032458.1; PID:G3449286
A:Experimental source: brain
C:Genetic8:
A:Gene: MEGF1
C:Superfamily: rat MEGF1 protein; cadherin repeat homology; EGF homology; laminin G repe
F:33798-3949/Domain: laminin G repeat homology <LGR>
F:33953-3985/Domain: EGF homology <EGF>
F:33992-4023/Domain: EGF homology <EGF1>

Query Match      8.2%; Score 1275; DB 2; Length 4351;
Best Local Similarity 24.0%; Pred. No. 2,8e-56;
Matches 534; Conservative 271; Mismatches 830; Indels 588; Gaps 77;
Qy      179  POFQPPSYQATVPENQAGTPVASIRAIIDPDEGAQLLETMDALFDSRSNQFSLDPT 238
Db      2377  PKFRPQYKAVANSEIATCGHLVLAQVADLPDIDGTSRLY--LILSGQDRHFSINSTS 2433
Qy      229  GAVTTAAELDETKSTHYFRVTAODHGMRRSALATLLIIVTDTNDHDPVEQCGYKSL 298
Db      2434  GLISFNLCKKQDSSYMLRGASD-GVFR--ATPVVITNTANKSPFQONVYAEEL 2490

```

```

Qy      299  RENLEVEYELTVRATDGDAPNNAILYRLLEGSGSGSPSEVEFIDPRSGVIRTRGPVDR 358
Db      2491  AENAKVGTKEVLEIADIDSGPYGVDTIINKLAG---EFFINFR-GQITTLQKDR 2546
Qy      359  EVESYQLTVEASDQGRDPFRSTTAAYFLSVEDNDNAPQSEKRYVQVVEDYTPGAPV 418
Db      2547  -NSTERYIAIKWARRDGGKVAFCVTKIILTDENDNAPQKASGYTVISNSVRSDPI 2604
Qy      419  LRYTASDCKSNMVAHYSIMS----- 440
Db      2605  IQVLAVIDDEGRNADVTVSDSTEDLAELIEVNPTGVVKVKSGLVLENRAVDENIKA 2664
Qy      441  ----- 440
Db      2665  QDGGPRHMDSLVPRLOVVRBEIPLPKSEBLYTPSAPEDLPESGTSQVAVAAQDPI 2724
Qy      441  -----GNARQPYLDAQALDVSPLDYETTKYTLRVA--ODGGRPLSN 486
Db      2725  YSLVQGTTPESNSDVSFLDDQDTGLKVRKAMDEHSTKMYQIDIMACHPHEDTLVSLVS 2784
Qy      487  VSGLVTVQVLDINNAPIFVSTPQATVLESVPGLVLYHQALDADAGNARLEYRLA- 545
Db      2785  VS-----IQVEDVNDNRPFVFEADPYKAPLTENNPGGTVVIQVTANDQDTGSDQVYSRLSV 2840
Qy      546  --GVGHDPPTINNNGTGMISVAELDEBEVDYFSGYEARDHG--TPALTASASVTVLD 602
Db      2841  EPGSNHHELFVAVDESCHITTLQELDCETQVTFYVVAIPHGQITQLSSQALVEVSLTD 2900
Qy      603  VNDNNPTPTQPEYTVRLNEDAAVGTSVYVSAVDRDA--HSVITYOITSQNTNRPESIT 659
Db      2901  ENDNPPRASEDYGSVVENNEPGLVATLKTLDADVSDQRQVTCVITTEGDPIGQFIS 2960
Qy      660  SQSGGGLVSLALPLDYLERQVYLAATASDGTRODTQIIVNVYTDANTHRVFPSSHVTV 719
Db      2961  QVEDEMRISRKTLDRHIAKYLLRYTASDGFQASVVEFVAVDINDNSQCSQLLYTG 3020
Qy      720  NVNEDRPAGTIVVLSATDEDTGENARITYFME--DSIPQRIADATGAVTQAELEDYDQ 778
Db      3021  KYREDVTPGHILKVAIDVDMQNAQITSLHSPGAQBFRLDHTGELTLYLVDNERK 3080
Qy      779  VSYTLAITARDNGIPQKSDTTYLEILVNDVNDNAPQLRDSYQSGVYEDVPEPTSLQIS 838
Db      3081  DVNMLVAKATDGG--GQSCQAEVTLHIEDVNDNAPREFPSCHDVAIVEDNTVTKPVAVF 3138
Qy      839  ATDRDSGLNGVFTPTQGGDGDGFVYESGIVRTLRRLRENVAGVYLAAYDKGM 898
Db      3139  ARDPDQANMQVVSILT--DSADQSFSDATSGVIRLEKPLQVRASSAVELTVRASDLGT 3196
Qy      899  P-PARTMEAVTVTVLDVNDNPPVEODEFVFEENSPIGLAVARV--TATDP-DEGTNAQ 955
Db      3197  PIPSTIGTVTVSVAGLEDVIFILANHEHQVEEDAPIDMEVHLATLTRPGEKSG-- 3254
Qy      956  IMYQIVENGINPEVFOQDIFSGELTALVDLYEDRPEYVLVIA--TSAPLVSRATYHVR 1012
Db      3255  --YHITGNEQCKRFLAHGTGLTVNGSLDPEETPKYFLGISCRKSSSSISDVTYIYIN 3312
Qy      1013  LLDENDNPP-----VLGNFELLFNMYNMRSSPFGAIGVPRANDP--ISDSLTVSEPR 1066
Db      3313  VTDVNEHHPFTHDLYTVRVLNNAV-----GDVILTVSASDDGQPVNSALTYSLVG 3364
Qy      1067  GNBLSLVILNASTGELKLSRALDNNRPLKAIMSVLSDG-----VHSYTAQCALVITYITD 1122
Db      3365  GNGGHTINPKKGLQVAKALDMEQRPYSILRARADSQGPLHEBT-EVAVEVDVND 3423
Qy      1123  EMLTHSITTLREDSPPRFPLSLGFLIQAVAATLATPPDHVVVFNVQRTDAFGHIILN 1182
Db      3424  -----NPRPFRQ-----LWYSTVQENSPIG--IKV 3447
Qy      1183  VSLSVGPQPGGGGPPPL-----PSBDLQRLYLNBSLTAISQORVLPFDNITCLAREPC 1237
Db      3448  LQTLIDDPDFQONPFYFRITRGNTGSVFRVTPDGMVLTASLIS-----KKAR 3496

```

```

QY 1238 ENYKCVSLRFPSSAFPIASSSVLPFRHPVGLRCRCP-----G 1279
DB 3497 EWMQIMTEV---SBSGLPLSSSTLVRQVTE---QGRYPSLPLEISITKBERGG 3550
QY 1280 FTGDYCTEVD---LCYSRCPGPHGCRSRGGTTLCDYGTGHECEVSANSRGCTPG 1335
DB 3551 MIGIHAITDDBPOTLTYS---LEOBGGIDRFTYTGASGK----- 3588
QY 1336 VCKKNGCTVNLVGVFGCDPSGDFEKPYCQVTRSPFASFTIRGLRQRPHPLTALS 1394
DB 3589 -----ILASQGLPHGRISFNTVS 3607
QY 1395 FATEKRDGLLNYGRFNEKIDFALVLYORQVOLTFSAGSSTTVSPFVGVSDGQMT 1454
DB 3608 -----DGTFTTTGVVHVHMEBVPVQAVMLGF-----HQLTEBELVSDHMRN 3652
QY 1455 VOLKTYKPLLGOTGLPQGSBQKAVVVTYDCCDTGVALRPGSVLGNYSQAQGTGSGK 1514
DB 3653 LQRFSLMLDVKRAMIHASLQPAEVTAGVD---VLTVF-----ERHSQTS 3695
QY 1515 KSIDLTPGLLGGVPLPESPPVNRQFVGQMRNLYQVDSHIDWADFIANNGTVPQCPAK 1574
DB 3696 YDLQELASALAHSHREIEHVGIRMR-----SALPVVPCQ 3730
QY 1575 KAVCDSTNCHNGGTGVNQMDAFSCCEPLGFGKSCAQEMANPOHFLGSSLVAMHGLSLP1 1634
DB 3731 GQSCQDQTCO-----ETVSLERVPVSYSFARLSITLPRHIGRN-CSCNGTTLAR 3780
QY 1635 S-----QPVTLISLMEFRTRQADVLLALITGRSTITVLQREG--HYMLSVEGT 1680
DB 3781 SGGSVQVYRPLEAQMWQIHFTYKTLQFWMALM-FTNETASISIKLANGSHLEYHCPG- 3837
QY 1681 GLQASSLRLPERGANDGDHMAOLALGASGAPGHAISFPYGOQBARAGNLPRLHGLHLS 1740
DB 3838 GFTGNLSRRP---VNDGQMT--SMLEBRDTSVHLVDI---TDNASLVPRECGHAKTE 3890
QY 1741 -NITVGI--PGPAGVAGAFRGCLQGVRS-DTPGCVSLDPSHG--ESINVEQCGSLP 1794
DB 3891 ROLLAGLVPNSPSSNVSILGFBEGCLDAVVVNGERLILGHEKMKMEGRLEFVALSQCMWP 3950
QY 1795 DPCDSNFCPANVSQSNWMS-YGSCSDPGYTGNC--TNVCOLNPEHQSVCTRPSAP 1850
DB 3951 TACSQSCNLNGSGSPALSGYLCCRPSPSGRNCBLAGRENTSAPCQBGCTCV---SSP 4007
QY 1851 HGVTCECPMYLAGPYCESTR-----IDPCPRGMWGH-----PTCG 1885
DB 4008 EGTSCNCPHYTGBRCMEARCGSGHCLTPEIKKQDWQOQERIVITVALPLVIATVG 4067
QY 1886 -----PCNCDVSKGPPDCKNTS-----GECHCKENHRP 1915
DB 4068 LLLYCRRRKSHKPVMTB-----DPDLARSIGVDTQASPAIIDLPLNTSSCN---MLNCP 4119
QY 1916 PGSPFCLCCPYTGLSLR-----VCDEBDQCP--KPGVIGROCDRCDNPAETTN 1967
DB 4120 BPSSTISVNELVTFPGSSKQPMVCSVPRLLPRANVSHPG-----HEPLIKTWS 4170
QY 1968 GCEVNDSCPRALFAGTW-----W-----PRTFGLPAAPCPKGS 2003
DB 4171 GEBELVYS-----GAAVMPRTYSKKKMEVPHPETMGTLPRSRHNVG-PAVMPDPTGL 4224
QY 2004 FG--TAVRHCDEHRGMLPRNLFN-----C--TSITF----- 2030
DB 4225 YGFPPLLENNKAPLRPRYSNQNLEDLMPRRPSPRBNILAPCLNEVYALISYHSQFR 4284
QY 2031 -----SELKGFARLORN-----BSGLDGSRSQOOLALLRNATGHTAGYFG 2071
DB 4285 QGGGGPCLAGGTYGVSMKLSRAGPSYADCEVNGSPATGHSQRA-----PRNYEG 4335
QY 2072 SDV 2074
DB 4336 SDM 4338

```

## RESULT 8

T20721

hypothetical protein F25F2.2 - Caenorhabditis elegans

C/Spectes: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 19-May-2000

C/Accession: T20721, T21343, T23842

R/Accession: R.

Submitted to the EMBL Data Library, August 1994

A/Reference number: Z19314

A/Accession: T20721

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;M12&gt;

A/Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone F10611

R/Accession: R.

Submitted to the EMBL Data Library, August 1994

A/Reference number: Z19410

A/Accession: T21343

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;M12&gt;

A/Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone F25F2

R/Accession: R.

Submitted to the EMBL Data Library, June 1994

A/Reference number: Z19806

A/Accession: T23842

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-4307 &lt;M13&gt;

A/Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2

A/Experimental source: clone M86

C/Accession: C.

A/Accession: CESP:F25F2.2

A/Map position: 3

A/Intons: 79/3; 167/3; 251/3; 328/3; 587/3; 625/3; 737/3; 1006/1; 1190/2; 1494/2; 1770/

C/Superfamily: rat MEGP1 protein, cadherin repeat homology; EGF homology; laminin G repe

F/761-3900/Domain: laminin G repeat homology &lt;LGR&gt;

Query Match 7.0%; Score 1092; DB 2; Length 4307;

Best Local Similarity 21.6%; Pred. No. 6, 4e-47;

Matches 417; Conservative 248; Mismatches 641; Indels 628; Gaps 58;

```

QY 232 FSLDPVGA-VTTAEE-----LDRRTKSTHVRV-----TAQDHGMRRSAL 272
DB 2410 FSLAPTSNSSTSKRAMVVALDRKTRGTHVSRVNLVHDXKQYQPLVED--ATGRRAF 2467
QY 273 ATTLIVTDNDHDPVFEQOEYKESLRNLEVGKVLTVRATDGAAPPANILYRLBGS 332
DB 2468 STLTLSVIDINDKPPFPVLPFYSTISBSAKGDTVMVWSATDD--ENDTIEYSLD-- 2523
QY 333 GSPSPVPEIDRSQVIRKRGVDRVDEEVESYQLTVEASQGDPPGRSTAAVPLSVEDD 392
DB 2524 -GSESQFVSHPRQGTIVAKCLEHAGVTLSTLKATDSANPPHATTVEV--NIASE 2580
QY 393 NDNAPOFSEKRYVVOVRDVTGAPVLKVTASDRDKSAAVHVSIGMSGNAGQ----FY 448
DB 2581 SVKVPFSSHLLFVMEADVGNVIGKQVETIDE---IRPFIIVPSBSDSFLPS 2637
QY 449 LDAQGALDVSPLDYETTKETTLAVRADGGRPPLSVSGI--VTVQVLINDNAPITV 506
DB 2638 VERSTGKILVKSLSLRERKQKMTIRADAA--GVAHTTVTVDIDGVNDNAPAH 2692
QY 507 STPFOATVLESVPLGVLYLVQAIADAGDNARLETRLAGVHDPFTINNCGMISVA 566
DB 2693 GYERTIIEBDAVGTSTIIFSAMDRDPSGRIRFSL--VEBNPYFDLNMESGMLTAS 2750
QY 567 ELDRREVDFYSGVARDHGPALTAASVSVTVLVDVNDNPTFTQPEY----- 615
DB 2751 QLDREKIDHYKLIARATBGG--GFNTDLPFTIVSDVNDSPQFKEKEBFNIDLHPSTSP 2808
QY 616 ----- 615

```

```

Db      2809 ILHFSIKODDLSPNNVSQFPIPKNEGVWIDSNNDLLKRPETIVENKQYQKLTARD 2868
Qy      616 -----TVRLNEDAAVGTSV-----VYTSAYDR 637
Db      2869 GVPETSTKVKINKSSKSDIRCPANKVTYLLAENSKGTIVLGBESSLLGPNTVPLSDN 2928
Qy      638 DAH-----SVITYQI-----647
Db      2929 DGNVFPVFNRRNGIVKYESDQIDYEKNQQLFHHRLTIQDNSEVCKELITVLEINLNDNRP 2988
Qy      648 -----TSGANTN-----RPSITSGSG-----GLVS 668
Db      2989 KLIKELKVISIDENLEPTSEDAQYITRIYADADPDEIKFRWVDGGLFQJIDINGVYT 3048
Qy      669 LALPLDYKLERQYVLAVTASDGTROPTAQIVNVNTDANTHRPVPQSHYTVANVERBPG 728
Db      3049 VVKPLDSEILGFPNLTVASDGEFEDKATILVTVDQNNAPTFEKSTYSMKWESBSIG 3108
Qy      729 TVVVLISATDEDTGENARITYEMEDS-IFQF-RIDADTGAVTTOAELEDDQVSYTLAT 786
Db      3109 YELAHFRASGCD--QNETIEYIAKPSDVTSFVNTAQTGILITAKPLDEFTLSALKLTIV 3166
Qy      787 ARDNQIPQKSDTFTYIEILVNDVNDNAQOFLRDSYGSYEDVPPTSVLQISATDRDGL 846
Db      3167 AADSGVPLETRAQVQVSVWDEBNDNAKFEKEKYGVKENSKEGKVLTVRALDVDSH 3226
Qy      847 NGRVFPPTGGDGDGD-----FIVSTSGIVTTLRLRBNVAQVLAAYVDKMPA 901
Db      3227 FPAVSTELTIVSEITTDTPVLPPAINS-NGDVLIQSSIDERYEIKKYNLVIYAKDGRPPL 3285
Qy      902 RPTMEVTVTVLVDNDNPVFEODEPVPVBEENSPIGLAVARVATPDDEGTNAQIMQYIV 961
Db      3286 REBALIEIHEBENBHAFTFDCCMTALVQEGEATHLIKSVSLDPPKGAFTVYI 3345
Qy      962 EBNIEVPFOLDIFSGELTALVDLYDRPEYLVIT-----QATSAPLVSAVYHRL 1014
Db      3346 QGDGAKSFYVN-EKLELTKAKLEYRKDKYLLTVAKOVKGTTDCPL-----TIFIRQT 3400
Qy      1015 DENDNPVLGNPEILFNNTVTRSSSPGGAGIYVANDPDIDS---LTSFEGNEIS 1071
Db      3401 SR-HAFTYKPKMLQINTL--QNELPFGITIGRLVASEDEBDONGILRFLGVBESIQS 3454
Qy      1072 -----LVLLNASTGEL-----KLSRALDNNRPLEAIVSVLVSGVSVTAQCAL 1115
Db      3455 PRAQVQESSTHLFRDPTVGTIWSHDSITQGLHT-----FNTVYDSKNTSYVAV 3507
Qy      1116 RVTITTDKMLTSTLTLEDMSPERPLSPLLGLFIQAVATATPPDHV-----VFN 1168
Db      3508 HTSINDVIDIAVSTRIRMSGVDEFRMKVKEFRRIISHHLNLDSSIQLSIVQAVPS 3567
Qy      1169 VORDDPAQGHILNLSVSGPFGPGGPPPLPSBLOGL-----YANRSL 1216
Db      3568 TISERSRNRNSMEDVEILMTAQRGLRG--YLPKPHIYRLKNDPQNNNDQSRMRYOI 3625
Qy      1217 TAISAQVLPFPDNLICLRBPC-----ENYKRCVSLRFPDSSAPPIASSVLPRIHV 1269
Db      3626 TEMCT-----TGVTLRGKREVEILIEDSWTKYSTDDFSFVSPRHSSA-----3669
Qy      1270 GGLRCCRPFGTADYCETEVDLCYSRPGPHRC--RSREGGYTCICRDGYTGEHCESA 1327
Db      3670 ---QCICPFGGKCEVETNOCAPCRQWOLCIPSVANSTYECVCPJGMEGDKCSVPS 3726
Qy      1328 RSGRCPTGVCKNGGTCV---NLLVGGFKDCPSGDEPKFYC--QVTRSPPAHSFIFRL 1383
Db      3727 ---CQNDGKCLBAELSVGG---DGFELISLSNIEIR-----3758
Qy      1384 RORFHETLALSPATKRDGLLYNGRFRNKHDFVALEVIOBOVOLTFSGBSTTVSPV 1443
Db      3759 ---MELIIBLKTTHNGIIMS---RGRDPTMLRLVNGTPEYMANAGTGTGIVTS--3808
Qy      1444 PGVSDGQHHYVQKTKYKPLGQGLPGPBBQKAVVTVVCGCTGVALR-FGSVLGNY 1502

```

```

Db      3809 KTSVVDGQWHLIATIS-----RRQRRTMTVDDEDLQEAFFSPISGTVINL 3852
Qy      1503 SCAAGTGGGSKKSIDLTLGPIILGGVPLPESFPVRKQPVGCERNLOYDSNHT-----D 1557
Db      3853 HRYSQLVLYGAK-----VDDGELTDGVS-----ACFRTISVDKMKVLTIRQG 3894
Qy      1558 MADFIANNNGVYGCBA-KKAVCDNSNTCHNGGTGVNQMDA-FSCGCPILGFGKSCAQEMAN 1615
Db      3895 MCLF-----GAQPGSALTSFPCNDLPQHAAGTICISQSHFRCEPCPSKSNVCEIDL--3948
Qy      1616 PQHFLGSSLVAMHGLSLPISQPWTYLSMFRTRQADGVLLQAITRGRSTTYLQLRBHWML 1675
Db      3949 -----3948
Qy      1676 SVETGLOASSURLEPERGANDGDWHQALALGASGPGHALLSPYGGQARAGNLGPRLLH 1735
Db      3949 -----3948
Qy      1736 GLHLNITVGGIPGPAGGVARGCLQGVRSVDTPEGVNSLDPEHGESINVEQCSLPD 1795
Db      3949 -----E 3949
Qy      1796 PCDSNPCPANSYCSNDWDSYSCSDPGYGDNC-----TNYCDLNPCEHGVCTRK 1846
Db      3950 FCASSPCPTGICQIPFYNDYLCCKPNGFTGKHCEARGFEDHETSCKNNVCGTSGQCTISI 4009
Qy      1847 PS---AHGTYCEC 1857
Db      4010 PHSLESSDFICNC 4023

```

## RESULT 9

T27110

Hypothetical protein Y52B11B.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27110

R:Renard, N.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20311

A:Accession: T27110

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1439 &lt;MUL&gt;

A:Cross-references: EMBL:AL032638; PDB:CAA21627.1; GSPDB:GN00019; CESP:Y52B11B.2

A:Experimental source: clone Y52B11B

C:Genetics:

A:Gene: CESP:Y52B11B.2

A:Map position: 1

A:introns: 11/2; 103/1; 183/2; 299/1; 406/3; 502/1; 585/3; 884/2; 969/1; 1017/1; 1185/3;

Query Match 5.3%; Score 817; DB 2; Length 1439;

Best Local Similarity 28.5%; Pred. No. 1.5e-33;

Matches 319; Conservative 173; Mismatches 474; Indels 154; Gaps 52;

```

Qy      176 NTAPOFQPSYQATVPENQAPTPVASLRATDPDEGAGRLBYTMDALPDRSNQFSLD 235
Db      272 DNVPTFRPLTYAQVREDIPINQTLIKTAVDKDTGDSKRTIYSVD-----NNEPSIN 324
Qy      236 PVTAQVTTAEELD-----REKSTHVFRTVTAQDHQP--RRSALATLTIVTDNDHPVF- 289
Db      325 S-NGSISAKVLLDADQLNHRFVYRVNVTADHGPVSLSSAMHIRTENTNBSAVFL 383
Qy      290 EQGEKESLEENLVEGYVTLVRAATDGAAPNANILVRLLEGSGSPSE--VFETDPSG 347
Db      384 PTSQITAFVADAGGTFVIOIQARDAD--RDEVTYSFMKNGRSYQKNLFSIDERTG 440
Qy      348 VIRTGFPV---DREAVES-YQITVEASDQGR---DPGRSTT-AAVFLSVEDDDNNAPOF 399
Db      441 LVKLRHGVSADLAEANPILVTIVQDDGSCCVYPSKTHSYATLLIGIDVNNKKEF 500
Qy      400 SE-KRY--VVOVRDV--TPGAPVLKRYTASDRDGSNAVHYST--MSGNARQGYLDAQ 452

```

Db 501 PDCAKYSDIAKIMEGTYKTDPTTIVKATDSDSANGDIVSLYYTOSSESRKAFVLRQ 560  
 Qy 453 TGAADVVP---LDYETKEVTLRVRAOQGGRPPLSNVSGLVVQVLDINDNAPVSTP 509  
 Db 561 TGVLT-TPSRVAVVDRETRPREVDYVKAITDRGDRPLIGFCOF-SVEVVDINDNSPQERPS 618  
 Qy 510 FQATVLESVPGLVLAHVQALDADAGDNALRYRL-----AGVGH-----DPPFTINGTG 560  
 Db 619 YETSVSRFEAVGTSVITVFAFDNDAAHNAITVSLSDITAGBEHQNDLDFELVNRASG 678  
 Qy 561 WISVAALDREEDVDFSGVEADHGT-ALTASASVSVLVLDVNNPFT-TOPEYT-- 616  
 Db 679 EITLIPITPKTKCFI-FNVIAADDNGIPALOSSAQVTLAVLDKQKAPKQMSPPCKKQ 737  
 Qy 617 VRLNEDAAVGSVVTASAVDR--DAHSVITYOITS-----GNTK-----NRPSTNSOG 663  
 Db 738 ITVDENVELAKVILRCRAVSSGDSRSDVYKLTASGGPKAKESKPEQKPF---ENG 793  
 Qy 664 GGLVSLAL--PLDYKEROVLAVTASDGT-----OD-----TAQIVAVV 702  
 Db 794 NENVEVVMIEGLDYGQNNVTTLTLTATVGPVPIQKKNVCFDFQDMTSRVASTKTEVEV 853  
 Qy 703 TDANTRPVPOSHYVNVNVE-----DRPAGTVVLISATDEPT-GENARITTFM-- 751  
 Db 854 RVNDVVPFTVDLFTGTIDEEMTPNEHLEKTKGKPIVYKALDTSDDGQNEVHYRIYG 913  
 Qy 752 ---EDSIPQRIADAGVTTQAEIDYEDQVSYTLAITARDNGI-----PQKSDTTY 800  
 Db 914 EANGETTKFRIDELTGEIFPNEKFPREKIDMTILVYKASDRSVSLPGANGPK-DNK 972  
 Qy 801 LELVNDVNDNAQPLRDSYQGSVYEDVPEFTSVLQISATRDSGLNGRVFTTPOGDDG 860  
 Db 973 VQIVINDVNDNAPFEEQKTIQVKESEGEHGVITIKAMDLDKHSNLR--YHLIGAGGG 1030  
 Qy 861 DGGFTYESTSGIVRTLRDLRENNVAVYLRAYAVDQGMPPARTPMVYTVLDVNDNPPV 920  
 Db 1031 RLPFGVRTSGTIFVKEPDPFASDQYHLVLASD--GRHNATF--NVYHIEBVNDNNAQ 1087  
 Qy 921 FEODEF-DYFVENSPGILAVAVTATDPDEG--TNAQIMQIVEGNIPEVFOIDISGEL 978  
 Db 1088 FBOQKATVIEEDVDIPIKVLFWVHATDADQDEKSRIVYRLGQADAVFRIGKSGSI 1147  
 Qy 979 TALVDLDYEDR---PEYVLYIQATSNP---LYSRATVHYRLDRNDPVLGNFEILFNN 1032  
 Db 1148 ELVKALDRPBPAGVPMNFVQALIDDGNGLVGAVOVVNDINNSPIFP--ERLFGY 1205  
 Qy 1033 YVTNRSSSPFGA-IGRVPAHDIDISLTYSEFG-----NELSVLNASGELTL 1084  
 Db 1206 IEBNRPISDGVYFMDVQARDFDPTTENANIEGIVRNKLINGSVFRIDQNGTKIRA 1265  
 Qy 1085 SPALDN--NRPLRAINSVLSD-GVHSAVT--AQCALRVTLITDEMLTHSITL---RLSDM 1136  
 Db 1266 MSLDIBESISERRPILEVANDGVSRBEPANVTIKATDMDNNAFPPEKTRIEGSEVET 1325  
 Qy 1137 SPERFLSPILGFIQAVATLA--TPPDHVVFNVRDTPD 1174  
 Db 1326 AP-----IGAAMSFSAFADDEAKDVFTQLSESD 1358  
 RESULT 10  
 T00043  
 BH-protocadherin-a - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
 C/Accession: T00043  
 R.Yoshida, K.  
 submitted to the EMBL Data Library, August 1997  
 A/Reference number: Z14075  
 A/Accession: T00043  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1069 <YOS>  
 A/Cross-references: EMBL:AB06758; NID:d1227200; PIDN:BA32597.1; PID:d1033562

C/Genetics:  
 A/Gene: Pcdh7  
 A/Map position: 5C3-D  
 Query Match 4.9%; Score 768.5; DB 2; Length 1069;  
 Best Local Similarity 28.1%; Pred. No. 2.8e-31;  
 Matches 224; Conservative 128; Mismatches 339; Indels 107; Gaps 17;  
 Qy 328 LLEGSG-----GSSVEFEIDPREGCVIRT--RGPVDRREVSXOLTVASDQGRD----- 375  
 Db 57 ITGSGEYVTSIESSEYKIDNLTGELSTERRIDREKLPQCOMIFDENECFLDPEVSV 116  
 Qy 376 PGPSTTAVF---LSVEDNDNAPQFSEKRVVQVEDVTEGAVLTVASDRDGSNA 432  
 Db 117 IGPSQSWDLFEGRIYVLDINDNTPFPSPVLTLLVENRPVGTLYLPLTADRPGNG 176  
 Qy 433 VTHYSIM-----SGNARGQTYLDAQTA---- 455  
 Db 177 IERYELQEPGGGGSGGRRLGPADSAPYPGGGNSASGGSGSKRLDAPEGGGGTS 236  
 Qy 456 -----LDVSPLDYETTKETLVRADGGRPPLSNVSGLV 492  
 Db 237 HSGRSVELQVADTPDSEKOPQLVKGALDRQSDYELTLRVGDGDPSPSS-QALDR 295  
 Qy 493 VQVLDINDNAPFVSTPQATVLESVPGLVLAHVQALDADAGDNALRYRLAGVGHDP 552  
 Db 296 VLTVDNNSPREFSVYEADLAENSAPGTPIQLRATVDLDVGVNGIIEVGAATESVR 355  
 Qy 553 --FTTNGTGMISVAALDREEDVDFSGVEARDHGTALTASASVSVTVLDVNDNPT- 609  
 Db 356 RLRLDRTSGMLSVLRIDREVNQRLFTMARQDGPCKTDKATVFNINDENNVPSI 415  
 Qy 610 -----FTQPEYVRLNEDAAVGSVTVASVADSD--AHSVITYQISGTRNRFSTS 660  
 Db 416 EIRTKGRPLPDGVANVADVLVTPIALVQVSDRDQSENGVYTTV-GDVPQLKAS 474  
 Qy 661 QSGGG-----LVSLALPLDYKLEQY---VLAVTASDGTQDTAQTIVNVTANTHRPV 711  
 Db 475 DREGQNKKKYFLHTSAPLDYETTEFNVVIAVDSGSPSLSNNSLVKVGDTINDNPPV 534  
 Qy 712 FQSHYTVNVEDRAGTVVLISATDEPTGENAITYFMEDSI-PQRIADTGAVTQ 770  
 Db 535 FQGSVEEYFPENNIIGRVATVLTADSGNMEIAISLSSVWGTALIDPSSGDIYN 594  
 Qy 771 AELDYEDQVSYTLAITARDNGIPQKSDTTYELIVNDVNDNAQPLRDSYQGSVYEDVP 830  
 Db 595 TVLDBEQDRLEFKNADKQIPVLQSGSTVIYQVADKNDPKRMQDVFTTYVENIQP 654  
 Qy 831 FTSVLQISATDRDSGLNGRVFTTPOGDDGDPFVSTSGIVRTLRLDRNNVAVYLR 890  
 Db 655 NSPVMVTVMDADKGRNAMESLYI---BENSNIFFIENDGTGIYSTWGFDRHQTYTFR 711  
 Qy 891 AYAVDQGMPPARTPMHVVTVLDVNDNPP--VFQDEEDVYBENSPIGLAVARTADP 948  
 Db 712 VKAVDQGPSPSATVATVSLFVMDENDNAPVYTLPPNNISYTLTPSSNRTVAATLATDS 771  
 Qy 949 DEGTAAQIMQIVEGNIEVEFQIDIFSGELTALVDLYEDREPEYVLYIQATSAPLVSAT 1008  
 Db 772 DDGINADLNTSGISVGNPKFLPEIDSTGVSGLVNGRLTKGHGIRLVVQVNDSGQPSGT 831  
 Qy 1009 ---VHRLDRNDNPPVL 1023  
 Db 832 TTLVHVFNESVSNATVI 849  
 RESULT 11  
 T00041  
 BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 01-Dec-2000  
 C/Accession: T00041; T00040  
 R.Yoshida, K.; Yoshitomo Nakagawa, K.; Seki, N.; Sasaki, M.; Sugano, S.  
 Genomics 49, 458-461, 1998

A::Title: Cloning, expression analysis, and chromosomal localization of BH-protocadherin1  
A::Reference number: Z14074; MUID:98277460; PMID:9615233  
A::Accession: T00041  
A::Status: preliminary; translated from GB/EMBL/DBJ  
A::Molecule type: mRNA  
A::Residues: 1-1072 <YOS>  
A::Cross-references: EMBL:AB006756; NID:g2979419; PIDN:BAA25195.1; PID:g2979420  
A::Experimental source: clone BH-Pcdh-b  
A::Accession: T00040  
A::Status: preliminary; translated from GB/EMBL/DBJ  
A::Molecule type: mRNA  
A::Residues: 1-1058, 'M', 1060, 'LH', 1063, 'Y', 1065, 'TVFG' <Y02>  
A::Cross-references: EMBL:AB006755; NID:g2979417; PIDN:BAA25194.1; PID:g2979418  
A::Experimental source: clone BH-Pcdh-a  
C::Genetics:  
A::Map position: 4p15

Query Match	4.9%;	Score 764.5;	DB 2;	Length 1072;
Best Local Similarity	27.2%;	Pred. No. 4.6e-31;		
Matches 231;	Conservative 138;	Mismatches 355;	Indels 125;	Gaps 21

QY	328	LLRSG-----GSPSEVELEIDPREGVLR.T--NGPVDREVESEYQJLWESADGRB-----	375
Db	57	IVTSGSEVTFSLBSESSEYKIDNLTGELSTSRRLIDREKLT.PCCQMIPBNEBCTLDFEVSIV	116
QY	376	FGPSTTRAAV---LSVEDDNDNAPQSEKRYVVOVEDVTFGAPVLRYTASDRDKSNA	432
Db	117	IGPQOSWDLPEGGQVTLVDINDNTPFPSPVLLTVEENRPVGTLYLLPTATDRPFGRRNG	176
QY	433	VWHSIM-----SOMAGQFPLDQNTA-----	455
Db	177	IERYELLQEPGGGSGGESRRRGAADSA.PYPPGGGNGASGGGSGSKRRLDLSBEGCGTN	236
QY	456	-----LDVBSPLDYETTKETYLRLRAQDGRPLPNSVGLYT	492
Db	237	PGRSSVPELVADTPDEKQPOLYVKALDBEQGDSYELTIRAVDGEDPSS--QALIR	295
QY	493	VQVLDINDNAPITFVSTPQATVLESVPGLVLYAOVADADAGNARLETRLAGVGHDP	552
Db	296	VLITVDNDNSPRPEKSYEADLAENSAGTPIQLRAADLDVGANGQLEYYGATGESVR	355
QY	553	--FTINGTGISVAELDREBVDPYSGVARHDGTALTBASVSTVLDVNDNNPT-	609
Db	356	RLRLDETSGMLSVLHRIIDREBVNDLRTVWARDDGQPPKTDKATVYINIKDENNVPSI	415
QY	610	-----FTQREYTRLNEDAAVGSVTVTSASVRD--ASHVITYQITSGNTRNFSITS	660
Db	416	EIRKIGRPLPDGVAANVAEDVLTPTLALVQVSDDOGENGVYTCV--GDVPOLKAS	474
QY	661	QSGGG-----LVSLAPLDYKLERQY---VLAVTASDGTRODTAQIVNVVTDANTHRPV	711
Db	475	DTEGDQNGKKYFLHSTPDTYEATREFNVVLYAVDGSGLSSKNSLTVKQDGTNDNPM	534
QY	712	QSGSHYVNVNVEDRPA GTTVVLISATBEDTGENAITYFMEDSIQ--FRIDATGAVTQ	770
Db	535	FGQSVETYYFPENNIPGERVATVLTATDADSGKNAEIAVSLDSSVWGIRFALIDPSGDIIVN	594
QY	771	AELVYEDVSTLAIATARDNGIPQKSDTTYELLVNDVNDNAPOLRLDSYQGSVEYEDR	830
Db	595	TVLREBQCDRIYFKVNAQDKGIPVLQGSITYVQVADKXNDPKKMOVPFTFYKXENLP	654
QY	831	FTSVLQISATRODGLNGRVFTFPQGGDGDGDFVSESTSGIVTRLRLDRENNVQVYL	890
Db	655	NSPFGVTVVMDADKGRNAMSGLYI---BENNNIISIENDTGLIYSTGMSFDRHQTTYYFR	711
QY	891	AYAVDKGMPRLAPTEVTVVYLDVNDNRP--VFQODEBDFVFEENSPIGLAVARYATDP	948
Db	712	VKAVDGGPSPRSATATVSLFVWDEBNDNAPYTLPLNISIYTLPLPSSNRYTVAATVLDSD	771
QY	949	DEGYNQAIQMYOIVESNIPEVFOLOIFSGELTALVLDVEDREBYLVYQATSAPLVSRAT	1008
Db	772	DDGINADLANYSIIVGGNPPKFLFIDPISGVSLVSKLTQNRHGHMLVVOVNDSGOPSOST	831

```

QY      1009 VHVALLDRNDNPVLGNFEILFNNYVNRSSFGGAIGRVPADP--DIDSITLYFE 1065
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      832 TTV-----VHVFNVSVENATNI--DSQIAR-SLHIFPLQDLADGPSYEIS 874

QY      1066 RGNELSLVL 1074
          :  :  :  :  :
Db      875 K-QRLSLVI 882

```

RESULT 12

A: Molecule type: mRNA  
A: Residues: 1-1467 <MNT>  
A: Cross-references: EMBL:AF111097; NID:G4185801; P1D:G4185802; PIDN:AAD09191.1  
C: Superfamily: alpha-latrotoxin receptor, calcium-independent  
C: Keywords: G protein-coupled receptor  
A: Status: preliminary; translated from GB/EMBL/DBD  
A: Title: The latrophilin family: multiply spliced G protein-coupled receptors with different  
A: Reference number: Z18869; MUID:99148828; PMID:10023961  
A: Accession: T18411  
A: Species: Bos primigenius taurus (cattle)  
C: Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C: Accession: T18411  
R: Matsushita, H.; Iellanova, V.G.; Ushkaryov, Y.A.  
FEBS Lett. 443, 348-352, 1999

Query Match	4.7%	Score 729.5	DB 2	Length 1467
Best Local Similarity	25.9%	Pred. No. 4.5e-29		
Matches 269	Conservative 139	Mismatches 378	Indels 251	Gaps 42

QY	1969	CEVNYDCSPALIEAGIMWPTTRFCFLPAAACEKPSFCSTAVAHCEHNG-WLP--PNIFNC	2025
Db	475	CE-----PBEYVR-VOMPATQCGMLVERCPCPKTRIGIASQCCLPALGIMNPRGDDLSNC	527
QY	2026	TSITFSELKGFABRLQNNESGDLDSGRSQOLALLRNATOHT-----AGYFGSDYKVAVOL	2086
Db	528	TSIPWNVQV--AQIKSGENA-----ANIASELABHTRGSIVAGDVSSVYKMEQL	575
QY	2081	-----ATRLLAHSTORGGLSTATQD-----VHFTENLLRVASALLDTPNK	2122
Db	576	LDLIDAOQLRLPTEBSACKNNKMKRRERTCKDYIKAVEITDNLRL-PEAL-----	628
QY	2122	RHMEILQOTEGG--TAWLQHYEAYASALAKNNRHATLSPFTIVTPNIVISVBLD-KGN	2178
Db	629	ESMKDMNATQATHTATMLDLVBEGALDLNDVREP--ARFLAKQNVLEVTYANTBEGQ	686
QY	2179	PAGAKLPREYALRGEOPPDETTLVILPESEVRETPTPVVRPAGPEAOPEELARQRRHP	2238
Db	687	VOELVFPO-----EYPS--ENSIOJSAHTIKQN-----	712
QY	2239	ELSGGRVAVSIITYRTLAGLPHNYD-----PDKSLRPKRPINTPVVSI	2285
Db	713	--SRNGVVKVFLIYNNLGLFJLSTENATVTLAGEAGSGGCGASL-----VANSOVIAA	764
QY	2286	SVHDDDEILPRALDKPYTVQFRLLETEBTRKPICVFNHNSILVSGTGMASARGEVVRN	2345
Db	765	SINKESRRV--FLMDPIFTVVAHLKAKNNHNNANCSFNWYSER--SMLGWSTQCGRLVBSN	821
QY	2346	ESHVSCQNMNTPSAVIMDVSRBENGEL--LPLKTLTYVALGVTLAALLTTFPFLTLARI	2403
Db	822	KHTTTCACSHLTFPAVIMARREIYQGRINELLISVTWVGIVISLVCAICISFTFCIRG	881
QY	2404	LRNSQHSIRBNTRLALGALVLELGINQADLPRACTVIALILHLFLTCTPSMALLBALH	2463
Db	882	LQDTRNTHNLCINIFLAELELFLVGDIKQYBIACIPFGLLHYFFLAAPSMWLCLEGVH	941
QY	2464	LYALTEVRDVNNGPMRFYMLCMGVAPFTGLAVAGDEPGYGNDPCMMLSIYDTLLIMSF	2523
Db	942	LYLLLVVEFSESRKTYVLLGVCPPALVVGIAALADHNSYGERKCMRLRVNDYFIMSF	1001
QY	2524	AGPVAFVANS--VELYTL--AABASCAORQGEKKGVSGLQPSFAVILLISATWLLA	2579

```

Db      1002 IGVSVFVIVNVLVFLMTLTHKRVNRSSEVLKPDSSRLDNIKSMALGAILFLGLTWARG 1061
Qy      2580 LLSVNSDTLLFHYLPATCNLCIOGPIFLSYVLSKEVRKAL-----KLACSRKSPDPA 2633
Db      1062 LFLINKESVMAVYLFTEFNAPQGVFLFHCALQKVKHKEYSKCLRHSTCCIRNP---Pq 1118
Qy      2634 LTTKSTLSSYNCSPSPYADG---RLYQPYGDS-----AGSLHST----- 2669
Db      1119 GAHGLSKTSAMRSNARYTGTQSRIRRMNDIVRKQTESFPMAGDINSTTLRGTMGNN 1178
Qy      2670 -----SRSGSGSPSYTF--LLREESALNPGQGPGLGDPGLFLESGDQDHPDPTD 2719
Db      1179 LLTNPVLQPRGTS-----PYNTLIAESVGFNP--SSPVPFNSPGS--YREPKHPLGSGREAC 1231
Qy      2720 SDSGLSLDDQSGSYASTHSSDSEEBEEBEBAAPGEGQWDSLGGPGERLP----- 2773
Db      1232 GMDTFLPLNGFNNSY--SLRSGD-----PPP-----GDGAPPPGRGNIA 1269
Qy      2774 -----LHSTPKDG-----GPGRKAPWPGDPFTTAKESGNGAPERRLRNG 2815
Db      1270 DAAPAEKMIISLVHNNLRGSSGAKGPPPEPVPVPVPGSGGEBAAG----- 1318
Qy      2816 DALSRBSGLPLPGSSAQP---HKGLKKKCLPTISEKSLRLPLEQCTGSSRGSSAS 2871
Db      1319 -----PGADRARIELLYKALEBPLLP--RAQSVLYQSDIDE-----SESCYAE 1360
Qy      2872 EGSRGGPPRPAPPROSL 2888
Db      1361 DGATSRPLSSPPGRDSL 1377

```

## RESULT 13

```

T18413
latrophilin-1, brain-specific - bovine
N/Alternate names: alpha-latrotoxin receptor, calcium-independent
C/Species: Bos primigenius taurus (cattle)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T18413
R/Matsuhashita, H.; Ieljanova, V.G.; Ushkaryov, Y.A.
FEBS Lett. 443, 348-352, 1999
A/Title: The latrophilin family: multiply applied G protein-coupled receptors with differ
A/Reference number: Z18869; M01D:99148828; PMID:10023961
A/Accession: T18413
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1472 <MAT>
A/Cross-references: EMBL:AF11098; NID:g4185803; PTD:g4185804; PIDN:AAD09192.1
A/Experimental source: Brain
A/Genetic: Lp11
C/Superfamily: alpha-latrotoxin receptor, calcium-independent
C/Keywords: G protein-coupled receptor

```

```

Query Match 4.7%; Score 729.5; DB 2; Length 1472;
Best Local Similarity 25.9%; Pred. No. 4.5e-29;
Matches 269; Conservative 139; Mismatches 378; Indels 251; Gaps 42;

Qy      1969 CEVVYDSCPRAIEAGIWPPTREGLPAAPCPKGSFETAVRHCDENRG--WLP--PMLFNC 2025
Db      480 CE-----PRRVR--VQMPATQGMVVERPCPKTRIGIAFOCLPALGLMNPSPDLSNC 532
Qy      2026 TSITSEILKGFARLQNRSGSLDSRQOLALLRNATQHT---AGYGSQVYKAYQL 2080
Db      533 TSPVNVGV---AQIKSGENA-----ANTASLARHTRGSIVAGDVSSVKKMEQL 580
Qy      2081 -----ATRLAHSTQRGSLATQD-----VHPEMLLRGSLALDTANK 2121
Db      581 LDILDAQLQLRPIERSAGNNVKKHRRCTCKDIKAVETVDNLR--PEAL----- 633
Qy      2122 RHWELIOOTGGG--TAMLQHYEAYASALQNMHTLSPFTIYTPVIVISVRLD--KGN 2178
Db      634 ESMWDMNATQAHATNTMLDVLEBGAFLADNVREP--AFLAKQNVVLEVIVLNTREQ 691

```

```

Qy      2179 FAGALIPRYEALRGQPPDLLETVILPESVELETPEPVVRPAGPGEAQBEEELARQRHHP 2238
Db      692 VQELVFPQ-----EYPS--ENSIGLSANTIKN----- 717
Qy      2239 ELISQEBASVAYIYTTLAGLPHNYD-----PDKRSIVPEKRPILNTPVYSI 2285
Db      718 --SRGNVAVVAFILVNNGLPLSTENATVKLAGESGPGGASL-----VNSQVIAA 769
Qy      2286 SVHDBELLPRALDQVTVQPRLLTEERTKPICYFMNHSILVSGTGSGMSACGVFBN 2345
Db      770 SINKESRY--FLMDPVLTVAHLAKHFNANCSFMYISER--SMLGTWSTQCGLYASN 826
Qy      2346 ESHVSCQCNHMTSFVAVMDVSRRENGEI--PLKLTLYVALGVTLAALLTFEFLTLRI 2403
Db      827 KHTTTCASHLNFVAVLAHREIYQGRINELLASITVWGVYLSVLCALICSTCPLRG 886
Qy      2404 LRSNQGIRNITAAVLAQVFLGINQADLPACTYAIALLHFLYICTYSWALLBLH 2463
Db      887 LQTDNNTIHKNCIMFLAELLFLVGIDKTOYEIACPIFAGLIHYFLAASMLCLEGVH 946
Qy      2464 LYRALTEVRDVNTGMRFYVYMGVPAFTGLAVGLDEGYNDFCMLSIYDTLIWSF 2523
Db      947 LYLIVVEFESRYSTRKTYTLAGYCPPALVGIATAIDRSYGTAKACMLRYDNTFINSF 1006
Qy      2524 AGPVAFAVMS--VFLYIL---AARASCAORQGEKGPVSGLOPSFAVLLLSATWLLA 2579
Db      1007 IGVSVFVIVNVLVFLMTLTHKRVNRSSEVLKPDSSRLDNIKSMALGAILFLGLTWARG 1066
Qy      2580 LLSVNSDTLLFHYLPATCNLCIOGPIFLSYVLSKEVRKAL-----KLACSRKSPDPA 2633
Db      1067 LFLINKESVMAVYLFTEFNAPQGVFLFHCALQKVKHKEYSKCLRHSTCCIRNP---Pq 1123
Qy      2634 LTTKSTLSSYNCSPSPYADG---RLYQPYGDS-----AGSLHST----- 2669
Db      1124 GAHGLSKTSAMRSNARYTGTQSRIRRMNDIVRKQTESFPMAGDINSTTLRGTMGNN 1183
Qy      2670 -----SRSGSGSPSYTF--LLREESALNPGQGPGLGDPGLFLESGDQDHPDPTD 2719
Db      1184 LLTNPVLQPRGTS-----PYNTLIAESVGFNP--SSPVPFNSPGS--YREPKHPLGSGREAC 1236
Qy      2720 SDSGLSLDDQSGSYASTHSSDSEEBEEBEBAAPGEGQWDSLGGPGERLP----- 2773
Db      1237 GMDTFLPLNGFNNSY--SLRSGD-----PPP-----GDGAPPPGRGNIA 1274
Qy      2774 -----LHSTPKDG-----GPGRKAPWPGDPFTTAKESGNGAPERRLRNG 2815
Db      1275 DAAPAEKMIISLVHNNLRGSSGAKGPPPEPVPVPVPGSGGEBAAG----- 1323
Qy      2816 DALSRBSGLPLPGSSAQP---HKGLKKKCLPTISEKSLRLPLEQCTGSSRGSSAS 2871
Db      1324 -----PGADRARIELLYKALEBPLLP--RAQSVLYQSDIDE-----SESCYAE 1365
Qy      2872 EGSRGGPPRPAPPROSL 2888
Db      1366 DGATSRPLSSPPGRDSL 1382

```

## RESULT 14

```

T17145
CLAB protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T17145
R/Sugita, S.; Ichtchenko, K.; Khvorochey, M.; Sudhof, T.C.
submitted to the EMBL Data Library, July 1998
A/Description: CL family.
A/Reference number: Z18712
A/Accession: T17145
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1510 <SIUG>
A/Cross-references: EMBL:AF081145; NID:g3695116; PID:g3695117; PIDN:AAC62651.1

```

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 4.7%; Score 728; DB 2; Length 1510;  
 Best Local Similarity 26.0%; Pred. No. 5.6e-29;  
 Matches 278; Conservative 133; Mismatches 387; Indels 270; Gaps 42;

QY 1969 CEVNTDSCRAIEMGIMWPRTRFGLPAAAPCPKSGFTGVHCHDRG-WLP--PNLFNC 2025  
 DB 475 CE-----PREVR-VQMPATQGMVERPCPKGTGASFOCLPALGLMNRGPDLSNC 527  
 QY 2026 TSITSEELKFAERLORNESGLDSGSQLALLRNATQHT-----AGYFSDYKAYQL 2080  
 DB 528 TSPWNVQ---AQIKSGEN-----ANISLARHTGSIYAGDVSSVTKMEQL 575  
 QY 2081 -----ATRLAHSTORFGLSATOD-----VHFTENLRVGSALLDTANK 2121  
 DB 576 LDILDAOLALRPIERESAGKNVNMKHKERTCKOYIKAVETVDNLK-PEAL----- 628  
 QY 2122 RHWELIQOTE--GGTAMLLQHTYEAVASALAQNKHHTYLSPTITVPIVIVSVRLDKGNF 2179  
 DB 629 ESMKQMNATEQVHTATMLDVLEEGAFLLADNVR-----EPARF 667  
 QY 2180 AGAKLPREYALRGEPDLETTVILPESVFRERTPPVVRPAPGGAOE---PEELARRQR 2236  
 DB 668 LAAK-----QNVLEVTVLSTE-----GQVQLVFPQEVYA--SES 700  
 QY 2237 HPELS-----QGEAVASVLIYRTLAGLPHNY-----DPDKSLRVPKR 2275  
 DB 701 SIQLSANTIKONSRNGVAVVFLVNNLGLFLSTENATVYKLAGAGTGGPGASL----- 755  
 QY 2276 PIINTPVVISVHDBELLPRALDKPYTVQFRLLSTEERTKPICVFNNHSILVSGTGWS 2335  
 DB 756 -VNSQVLAASINKSSRV--FLMDPVITVAHLEAKHFNANCSFNNYSER-SMLGWS 811  
 QY 2336 ARGEVFRNESHVSCQNMHTSPAVLMDVSRNGEI--LPLKTLTYVALGVTALALL 2393  
 DB 812 TQGCFLVBSNKTHTTACSHLTNFAVMAHREIYQGRINELLSVITWGVISLVCLAI 871  
 QY 2394 TFFPFLTLIRLSNOHGIRNLTAAAGLAQVLGLINQADLPACTVIALILHFLYCT 2453  
 DB 872 CISTCFPLRGLOTDRTNTHKNCINLFLABELFLVGDIDKTYEVAACP1FAGLHFFFLAA 931  
 QY 2454 PSWALLEALHLYRALTEVDVNTGPMRFTYMLGKGVPAFTGLAVGLDEGYGNPDCWL 2513  
 DB 932 FSWMLLEGVHLYLLVEVESEYSRTKYTYLGVCYCPALVGIAMADIRSGTAKACWL 991  
 QY 2514 SIYDTLIWSPAGVAVAVSMS-VFLYTL--AARASCAARQGFKKGPVSGLOPSPAVL 2569  
 DB 992 RVDNYFIWSPFGVSVIYVNLVFLMWTLHKMIRSSSVLKPDSSRLDNIKSMALGAILL 1051  
 QY 2570 LLLSATWLLALSVNSDTLLFHYLPATCNCIOGPFIFLSVYLSKVRKAL-----KLA 2623  
 DB 1052 FLGLGTMAFGLLPIKESVYMAVYLFITTPAFOGVIPIVPHCALQKKVHKYKICLRHVC 1111  
 QY 2624 CSKRSPPDALTTKSTLTSSYNCPSPYADRLYQPYGDSAGLSHSTNSGKSQPSYIPL 2683  
 DB 1112 CTRSP---FGAHGSLKTSKMSRNTRYTYGTQVPGG--RHIIHVSJLGRGR-SALP-- 1162  
 QY 2684 LREESALNCGGPGGLGDP-----GSLFLBEGDDQDHDVDTSD 2721  
 DB 1163 ---BSQKDPG-QQSGPBDPLTGLCPSRIRRMWMDTVRKQTESSPAG-----DIN 1209  
 QY 2722 SGLSLEDQSGSVASTHSSDEEBEEREAAP-----GEOGMDSL----- 2765  
 DB 1210 STPLINRGTMGNHLLTN-----PVLQGRGTSYNTLLASGVFNPS 1252  
 QY 2766 -----GPGKERPLIPLSTPDGPGPKAPWPDFFGTAKSS-----GNGAPE-ERLKEN 2814  
 DB 1253 PVPFNSPGSYREPKPFLGREGACMDTTLPLNGFNNSYSLRSGDPPPGDGGEPBRGRNTL 1312  
 QY 2815 GGLSRE-----GSLGPLPGSSAOPHKGILKKC-LPTISSEKSL--- 2853  
 DB 1313 ADAAAFKKIITSELVHNNLRGASGAKGPPPPPPVPVPGVSEDAAGPBGADRAIELL 1372

QY 2854 -----LRLPLEQCT-----GSSRGSSASRGSGCPPPRPGSL 2888  
 DB 1373 YKALEEPLILRPAQSVLYQSDLBESCTADGARSPLSSPPGDSL 1420

## RESULT 15

R17156

CLIBB protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T17156

R:Sugita, S.; Ichchenko, K.; Khavochey, M.; Sudhof, T. C.

A:Submitted to the EMBL Data Library, July 1998

A:Description: Cl family.

A:Reference number: Z18712

A:Accession: T17156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1515 &lt;STUG&gt;

A:Cross-references: EMBL:AF081147; NID:G3695120; PID:G3695121; PIDN:AA62653.1

C:Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 4.7%; Score 728; DB 2; Length 1515;  
 Best Local Similarity 26.0%; Pred. No. 5.6e-29;  
 Matches 278; Conservative 133; Mismatches 387; Indels 270; Gaps 42;

QY 1969 CEVNTDSCRAIEMGIMWPRTRFGLPAAAPCPKSGFTGVHCHDRG-WLP--PNLFNC 2025  
 DB 480 CE-----PREVR-VQMPATQGMVERPCPKGTGASFOCLPALGLMNRGPDLSNC 532  
 QY 2026 TSITSEELKFAERLORNESGLDSGSQLALLRNATQHT-----AGYFSDYKAYQL 2080  
 DB 533 TSPWNVQ---AQIKSGEN-----ANISLARHTGSIYAGDVSSVTKMEQL 580  
 QY 2081 -----ATRLAHSTORFGLSATOD-----VHFTENLRVGSALLDTANK 2121  
 DB 581 LDILDAOLALRPIERESAGKNVNMKHKERTCKOYIKAVETVDNLK-PEAL----- 633  
 QY 2122 RHWELIQOTE--GGTAMLLQHTYEAVASALAQNKHHTYLSPTITVPIVIVSVRLDKGNF 2179  
 DB 634 ESMKQMNATEQVHTATMLDVLEEGAFLLADNVR-----EPARF 672  
 QY 2180 AGAKLPREYALRGEPDLETTVILPESVFRERTPPVVRPAPGGAOE---PEELARRQR 2236  
 DB 673 LAAK-----QNVLEVTVLSTE-----GQVQLVFPQEVYA--SES 705  
 QY 2237 HPELS-----QGEAVASVLIYRTLAGLPHNY-----DPDKSLRVPKR 2275  
 DB 706 SIQLSANTIKONSRNGVAVVFLVNNLGLFLSTENATVYKLAGAGTGGPGASL----- 760  
 QY 2276 PIINTPVVISVHDBELLPRALDKPYTVQFRLLSTEERTKPICVFNNHSILVSGTGWS 2335  
 DB 761 -VNSQVLAASINKSSRV--FLMDPVITVAHLEAKHFNANCSFNNYSER-SMLGWS 816  
 QY 2336 ARGEVFRNESHVSCQNMHTSPAVLMDVSRNGEI--LPLKTLTYVALGVTALALL 2393  
 DB 817 TQGCFLVBSNKTHTTACSHLTNFAVMAHREIYQGRINELLSVITWGVISLVCLAI 876  
 QY 2394 TFFPFLTLIRLSNOHGIRNLTAAAGLAQVLGLINQADLPACTVIALILHFLYCT 2453  
 DB 877 CISTCFPLRGLOTDRTNTHKNCINLFLABELFLVGDIDKTYEVAACP1FAGLHFFFLAA 936  
 QY 2454 PSWALLEALHLYRALTEVDVNTGPMRFTYMLGKGVPAFTGLAVGLDEGYGNPDCWL 2513  
 DB 937 FSWMLLEGVHLYLLVEVESEYSRTKYTYLGVCYCPALVGIAMADIRSGTAKACWL 996  
 QY 2514 SIYDTLIWSPAGVAVAVSMS-VFLYTL--AARASCAARQGFKKGPVSGLOPSPAVL 2569  
 DB 997 RVDNYFIWSPFGVSVIYVNLVFLMWTLHKMIRSSSVLKPDSSRLDNIKSMALGAILL 1056  
 QY 2570 LLLSATWLLALSVNSDTLLFHYLPATCNCIOGPFIFLSVYLSKVRKAL-----KLA 2623



Db 1057 FLGLTWAFGLLPINKESVMAYLFTTNAFCGVIFVHFICALQKXVHKRYSKCLRHASYC 1116  
 QY 2624 CSKRPSFDPALTTKSTLTSSYNCPYADGRLYOPYGDSAGSLHSTSRSGSQPSY1PFL 2683  
 Db 1117 CIRSP---PGAHGSLKTSAMRSVTRYTGTVGOG---RHITHVSLGPRGR-SALP-- 1167  
 QY 2684 LREESALNPGQPPGLGDP-----GSLFLEGQDQHDPTDSD 2721  
 Db 1168 ---BQKDPG-GQSGPDPDLTFGLCPGRIRRMNDTVRKQTESFMAQ-----DIN 1214  
 QY 2722 SDSLBDQSGSYASTHSDSSEEEEEEAAFP-----GEGMDSL----- 2765  
 Db 1215 STPTLNGTGNHLLTN-----PVLQPRGTSPTVTLIABSVGNPSS 1257  
 QY 2766 ---GQAEPLPLHSTPKDGGPGKAPWPGDGTAKES---GNGAPE-ERLREN 2814  
 Db 1258 PVTNSPGSYREPKHPIAGREACGMDTLPLNGVFNNSYSLRSGDPPGDDGPPPRGRNL 1317  
 QY 2815 GDALSRE-----GSLGPLGSSAQPHKGLKKC-LPTISEKSL--- 2853  
 Db 1318 ADAAAEKMTISLHVNNLRGASGAGAKPPPPVPPVPGVSBDEAGPGGADRAABIELL 1377  
 QY 2854 -----LRLPLECT-----GSSRGSASBSRGSGPPPPROSL 2888  
 Db 1378 YKALEEPLLPRAQSVLYQSDLDESECTAEDGATSRPLSSPPGRDSL 1425

Search completed: February 11, 2004, 15:57:55  
 Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:51:34 ; Search time 25 seconds  
(without alignments)

5498.362 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545  
Sequence: 1 MRSFATGVLPPTPPPLLL.....ACTVDESSGSEFLFFNLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	15545	100.0	2923 1	CLR2_HUMAN
2	14657.5	94.3	2920 1	CLR2_MOUSE
3	11097	71.4	2144 1	CLR2_RAT
4	8974.5	57.7	3034 1	CLR1_MOUSE
5	8754	56.3	3014 1	CLR1_HUMAN
6	7833.5	50.4	3312 1	CLR3_HUMAN
7	7780.5	50.1	3313 1	CLR3_RAT
8	7732.5	48.7	3301 1	CLR3_MOUSE
9	4817.5	31.0	3579 1	STRN_DROME
10	1534.5	9.9	5147 1	FAT1_DROME
11	1516.5	9.8	4590 1	FAT2_DROME
12	1477.5	9.5	4705 1	CA2N_DROME
13	1419.5	9.1	3097 1	CA2N_MOUSE
14	1320.5	8.5	4349 1	FAT2_HUMAN
15	1275	8.2	4351 1	FAT2_RAT
16	1262.5	8.1	3298 1	PC16_HUMAN
17	1218.5	7.8	3317 1	CA2N_RAT
18	1217.5	7.8	3354 1	CA2N_HUMAN
19	1214	7.8	3354 1	CA2N_MOUSE
20	1207	7.8	3503 1	DS_DROME
21	985	6.3	2215 1	CD82_DROME
22	905	5.8	1507 1	CD87_DROME
23	858	5.5	1507 1	CA2E_DROME
24	853	5.5	1955 1	PC15_HUMAN
25	846	5.4	1026 1	PC15_HUMAN
26	833.5	5.4	1943 1	PC15_MOUSE
27	764.5	4.9	1063 1	PC17_HUMAN
28	710.5	4.6	1584 1	BA11_HUMAN
29	687.5	4.4	3343 1	YOC7_CAEEL
30	678.5	4.4	1007 1	CHC2_HUMAN
31	668.5	4.3	2240 1	CD89_DROME
32	667.5	4.3	793 1	CD87_HUMAN
33	667	4.3	936 1	CD45_HUMAN

34	662.	4.3	798 1	CD8E_HUMAN	Q9Y5e9 homo sapien
35	654	4.2	795 1	CD8C_HUMAN	Q9Y5f1 homo sapien
36	653	4.2	1522 1	BA13_HUMAN	Q60242 homo sapien
37	649.5	4.2	950 1	CD43_HUMAN	Q9Y5h8 homo sapien
38	648.5	4.2	1180 1	PC12_MOUSE	Q55134 mus musculu
39	647.5	4.2	950 1	CD41_HUMAN	Q9Y5f0 homo sapien
40	644.5	4.1	798 1	CD8D_HUMAN	Q9Y5f0 homo sapien
41	642	4.1	949 1	CD8E_HUMAN	Q9Y5f0 homo sapien
42	632	4.1	807 1	CD8B_HUMAN	Q9un66 homo sapien
43	630	4.1	941 1	CD8A_HUMAN	Q9un74 homo sapien
44	628.5	4.0	795 1	CD84_HUMAN	Q9Y5e5 homo sapien
45	628.5	4.0	797 1	CD83_RAT	Q63418 rattus norv

## ALIGNMENTS

RESULT 1  
ID CLR2\_HUMAN STANDARD; PRT; 2923 AA.  
AC O9HCT4; Q92566;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cadherin BGF LAG seven-pass G-type receptor 2 precursor (Epidermal growth factor-like 2) (Multiple epidermal growth factor-like domains 3) (Flamingo 1).  
DE CHER2 OR CDH10 OR EGFL2 OR MEGF3 OR KIAA0279.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20363102; PubMed=10907856;  
RA Vincent J.B., Skang J., Scherer S.W.;  
RT "The human homologue of flamingo, EGFL2, encodes a brain-expressed large cadherin-like protein with epidermal growth factor-like domains, and maps to chromosome 1p13.3-p21.1.";  
RT DNA Ref. 7:233-235(2000).  
RL [2]  
RN SEQUENCE OF 516-2923 FROM N.A.  
RP TISSUE=Brain;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain.";  
RT DNA Ref. 3:321-329(1996).  
RL [1]  
RN FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
CC - SUBCELLULAR LOCATION: Integral membrane protein.  
CC - TISSUE SPECIFICITY: Highest expression in brain and testis.  
CC - SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC - SIMILARITY: Contains 9 cadherin domains.  
CC - SIMILARITY: Contains 8 BGF-like domains.  
CC - SIMILARITY: Contains 2 laminin G-like domains.  
CC - SIMILARITY: Contains 1 GFS domain.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC EMBL: AF234887; AAC00080.1; -  
CC EMBL: D87469; BA113407.1; -  
CC HSSP: P15116; INCT  
DR Genem; HGNC:3231; CELSR2.  
DR MIM: 604265; -

DR GO; GO:0016021; C:integral to membrane; NAS.  
 DR GO; GO:0004930; F:G-protein coupled receptor activity; NAS.  
 DR Interpro; IPR000152; Aex hydroxyl.  
 DR Interpro; IPR002126; Cadherin.  
 DR Interpro; IPR000742; EGF 2.  
 DR Interpro; IPR001881; EGF\_Ca.  
 DR Interpro; IPR006209; EGF-like.  
 DR Interpro; IPR000832; GPCR secretin.  
 DR Interpro; IPR001879; hormn receptor.  
 DR Interpro; IPR002049; laminin EGF.  
 DR Interpro; IPR001791; laminin\_G.  
 DR Interpro; IPR00203; PLO\_cys-rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00053; laminin\_EGF; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF/LAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00179; EGF\_Ca; 1.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00282; Hormr; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 7.  
 DR PROSITE; PS00268; CADHERIN\_2; 9.  
 DR PROSITE; PS00221; GPS; 1.  
 DR PROSITE; PS00022; EGF; 1; 6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00649; G-PROTEIN RECP\_F2\_1; FALSE NEG.  
 DR PROSITE; PS00650; G-PROTEIN RECP\_F2\_2; FALSE NEG.  
 DR PROSITE; PS00227; G-PROTEIN RECP\_F3\_1.  
 DR PROSITE; PS00261; G-PROTEIN RECP\_F4\_1.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ TYPE EGF; 1.  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal.  
 FT CHAIN 32 2923  
 FT SIGNAL 31  
 FT DOMAIN 32 2380  
 FT TRANSSEM 2381 2401  
 FT TRANSSEM 2402 2416  
 FT TRANSSEM 2417 2437  
 FT TRANSSEM 2438 2438  
 FT TRANSSEM 2439 2459  
 FT TRANSSEM 2460 2480  
 FT TRANSSEM 2481 2501  
 FT TRANSSEM 2502 2519  
 FT TRANSSEM 2520 2540  
 FT TRANSSEM 2541 2560  
 FT TRANSSEM 2561 2581  
 FT TRANSSEM 2582 2591  
 FT TRANSSEM 2592 2612  
 FT TRANSSEM 2613 2923  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 505  
 FT DOMAIN 506 610  
 FT DOMAIN 611 712  
 FT DOMAIN 713 815  
 FT DOMAIN 816 921  
 FT DOMAIN 922 1023  
 FT DOMAIN 1028 1146  
 FT DOMAIN 1228 1286  
 FT DOMAIN 1288 1324  
 FT DOMAIN 1328 1366  
 FT DOMAIN 1367 1571

FT DOMAIN 1574 1610 EGF-LIKE 4, CALCIUM-BINDING.  
 FT DOMAIN 1614 1791 LAMININ G-LIKE 2.  
 FT DOMAIN 1793 1828 EGF-LIKE 5, CALCIUM-BINDING.  
 FT DOMAIN 1829 1867 EGF-LIKE 6, CALCIUM-BINDING.  
 FT DOMAIN 1883 1922 EGF-LIKE 7, CALCIUM-BINDING.  
 FT DOMAIN 1923 1955 EGF-LIKE 8, CALCIUM-BINDING.  
 FT DOMAIN 2316 2368 GPS.  
 FT DOMAIN 2743 2752 POLY-GLU.  
 FT DISULFID 1232 1243 BY SIMILARITY.  
 FT DISULFID 1237 1274 BY SIMILARITY.  
 FT DISULFID 1276 1285 BY SIMILARITY.  
 FT DISULFID 1292 1303 BY SIMILARITY.  
 FT DISULFID 1297 1312 BY SIMILARITY.  
 FT DISULFID 1314 1323 BY SIMILARITY.  
 FT DISULFID 1332 1343 BY SIMILARITY.  
 FT DISULFID 1337 1353 BY SIMILARITY.  
 FT DISULFID 1355 1365 BY SIMILARITY.  
 FT DISULFID 1378 1389 BY SIMILARITY.  
 FT DISULFID 1578 1589 BY SIMILARITY.  
 FT DISULFID 1583 1598 BY SIMILARITY.  
 FT DISULFID 1600 1609 BY SIMILARITY.  
 FT DISULFID 1797 1808 BY SIMILARITY.  
 FT DISULFID 1802 1817 BY SIMILARITY.  
 FT DISULFID 1819 1828 BY SIMILARITY.  
 FT DISULFID 1832 1843 BY SIMILARITY.  
 FT DISULFID 1837 1855 BY SIMILARITY.  
 FT DISULFID 1857 1866 BY SIMILARITY.  
 FT DISULFID 1887 1899 BY SIMILARITY.  
 FT DISULFID 1889 1906 BY SIMILARITY.  
 FT DISULFID 1908 1921 BY SIMILARITY.  
 FT DISULFID 1924 1936 BY SIMILARITY.  
 FT DISULFID 1926 1943 BY SIMILARITY.  
 FT DISULFID 1945 1954 BY SIMILARITY.  
 FT MOD RES 1591 1591 BY SIMILARITY.  
 FT MOD RES 1810 1810 BY SIMILARITY.  
 FT CARBOHYD 486 486 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1036 1036 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1182 1182 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1501 1501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1565 1565 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1741 1741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1827 1827 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1900 1900 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2024 2024 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2043 2043 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2061 2061 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2323 2323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2345 2345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 2923 AA; 317447 MW; 382757D915158BD8 CRC64;

Query Match 100.0%; Score 15545; DB 1; Length 2923;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPPATGVLPPTPPPLLLLLLLPPLLAGOVGCRSLASRGSGAGCAPMGLCPSS 60  
 |||  
 DB 1 MSPAPGVLPPTPPPLLLLLLLPPLLAGOVGCRSLASRGSGAGCAPMGLCPSS 60  
 |||  
 QY 61 SASNLMLYTSRCDAGTELTGHLVPHHDGLRWCPESBAHPLPPAPBCCPMSCRLIGIG 120  
 |||  
 DB 61 SASNLMLYTSRCDAGTELTGHLVPHHDGLRWCPESBAHPLPPAPBCCPMSCRLIGIG 120  
 |||  
 QY 121 GHLSPPGKLTLPBEPHCLKAPRLRCQSCGLAAPGLRAEBSPEBSLGGRRRNVTAAQ 180  
 |||  
 DB 121 GHLSPPGKLTLPBEPHCLKAPRLRCQSCGLAAPGLRAEBSPEBSLGGRRRNVTAAQ 180  
 |||  
 QY 181 POPPSQATVPENOPAGTVPASIRATIDPEGEAGRLRYTMDFDPSRSNQFSLDPVTGA 240  
 |||  
 DB 181 POPPSQATVPENOPAGTVPASIRATIDPEGEAGRLRYTMDFDPSRSNQFSLDPVTGA 240  
 |||

241 VTTAEELREKTSHTVPRVTAQDHGMPRSSALATITLVTOTNDNDPVPEOQEKESIRE 300  
241 VTTAEELREKTSHTVPRVTAQDHGMPRSSALATITLVTOTNDNDPVPEOQEKESIRE 300  
301 NLEVEGYVLTARATGDAPPNANILYRLLBGSQSPSEYFEIDPSSGYIARTGRAPDREB 360  
301 NLEVEGYVLTARATGDAPPNANILYRLLBGSQSPSEYFEIDPSSGYIARTGRAPDREB 360  
361 ESYQULTVEASDQGRDPGRSTTAAVFLSVEDDNDNAPQSEKRYVVOYREBDVTPCAPYL 420  
361 ESYQULTVEASDQGRDPGRSTTAAVFLSVEDDNDNAPQSEKRYVVOYREBDVTPCAPYL 420  
421 VTSARDKGSNAVHYYSIMSGNARQFYLDAGTALDVVSPLDYETTKRYTLRYAQAQGG 480  
421 VTSARDKGSNAVHYYSIMSGNARQFYLDAGTALDVVSPLDYETTKRYTLRYAQAQGG 480  
481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESYPLGYLVHVAOAIADADGNARL 540  
481 RPLSNVSGLVTVQVLDINDNAPIFVSTPFOATVLESYPLGYLVHVAOAIADADGNARL 540  
541 EYRLAGVGHDPFTTINNGTGMISVAEELDREBVDYFSFGVEARDHGTALFASASVTV 600  
541 EYRLAGVGHDPFTTINNGTGMISVAEELDREBVDYFSFGVEARDHGTALFASASVTV 600  
601 LDVNDNNFTPOPEXTVRLNEDAAVGSVWYVSAIDRDAHSITTYQITSGATRRKFSITS 660  
601 LDVNDNNFTPOPEXTVRLNEDAAVGSVWYVSAIDRDAHSITTYQITSGATRRKFSITS 660  
661 OSQGGVSLALPLDYKLERQVYLVASDGTREQDTAQIVNVTADNTHRPVFOSSHITVN 720  
661 OSQGGVSLALPLDYKLERQVYLVASDGTREQDTAQIVNVTADNTHRPVFOSSHITVN 720  
721 VNEBRPAGTIVLISATBEDTGENARITYFMEDSIPOFRIDADGAVTTOAELDEYDVS 780  
721 VNEBRPAGTIVLISATBEDTGENARITYFMEDSIPOFRIDADGAVTTOAELDEYDVS 780  
781 YTLAITADNNGIPOKSDPTTYLEIIVNDVNDNAPOGLRBSYQGSYVEDPPTSVULQISAT 840  
781 YTLAITADNNGIPOKSDPTTYLEIIVNDVNDNAPOGLRBSYQGSYVEDPPTSVULQISAT 840  
841 DRDSGLNRRVYTTPOGSDGDDGFIVESTSGIVRTLRRLREBNVQYVLRAYAVDKMP 900  
841 DRDSGLNRRVYTTPOGSDGDDGFIVESTSGIVRTLRRLREBNVQYVLRAYAVDKMP 900  
901 ARTPEVTVTVLDVNDNPPVFBODEFVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
901 ARTPEVTVTVLDVNDNPPVFBODEFVFEENSPIGLAVARVATDDEGTNAQIMYQI 960  
961 VEGNIPVFOJDIPEGELTALVDLYEDRPREVULYQATSAPLYSRATVHRLLDNRNP 1020  
961 VEGNIPVFOJDIPEGELTALVDLYEDRPREVULYQATSAPLYSRATVHRLLDNRNP 1020  
1021 PVLGNFELLFNNYVNNRSSPPGAGIGRAYPAHDPIIDSLTYSPFERGKLSLVILLNASTG 1080  
1021 PVLGNFELLFNNYVNNRSSPPGAGIGRAYPAHDPIIDSLTYSPFERGKLSLVILLNASTG 1080  
1081 ELKLSRALDNNRPLEAIVSVSDGHSVTAQCALRTIITDEMUTHSITLRLBMSER 1140  
1081 ELKLSRALDNNRPLEAIVSVSDGHSVTAQCALRTIITDEMUTHSITLRLBMSER 1140  
1141 FLSPILGLFIOAVATLATPBDHVVVENVORDTAPRGHILNLSVQCPRGPGGPF 1200  
1141 FLSPILGLFIOAVATLATPBDHVVVENVORDTAPRGHILNLSVQCPRGPGGPF 1200  
1201 PSEBLOERLYLNRSLLTAISAQVLPFDNICTLRPCENYMRCSVLFPSDAPFIASS 1260  
1201 PSEBLOERLYLNRSLLTAISAQVLPFDNICTLRPCENYMRCSVLFPSDAPFIASS 1260  
1261 VLFPRPIHVGGLRCPCPGFTGDIYETEVULCYSPCGPHRCRSREGGYCLCRDGYTG 1320  
1261 VLFPRPIHVGGLRCPCPGFTGDIYETEVULCYSPCGPHRCRSREGGYCLCRDGYTG 1320  
1321 EHCEVSASRGCTPGVCNKGCTCVNLLVGFKCPCPSGDFEKPQYQVTTSPFAHSFTTF 1380

1321 EHCEVSASRGCTPGVCNKGCTCVNLLVGFKCPCPSGDFEKPQYQVTTSPFAHSFTTF 1380  
1381 RGLRORHFTLALSFATKRDGLLXNGFRFKKDPALEAVIOGVOVLTFSAGSTTVS 1440  
1381 RGLRORHFTLALSFATKRDGLLXNGFRFKKDPALEAVIOGVOVLTFSAGSTTVS 1440  
1441 PFPBGVSDGQWHTVOLKTYNKPDLAQGTGLPOGSEOKVAVVTVDGDGVALARFGSVLG 1500  
1441 PFPBGVSDGQWHTVOLKTYNKPDLAQGTGLPOGSEOKVAVVTVDGDGVALARFGSVLG 1500  
1501 NYSQAQGTQGSKSLDITGPLLGGVPLPESFPVRMQPVCCRNLOVDSHIDMAD 1560  
1501 NYSQAQGTQGSKSLDITGPLLGGVPLPESFPVRMQPVCCRNLOVDSHIDMAD 1560  
1561 PIANGTVPGCPAKNVCDSNTCHNGGTCVNOMAPFSCBCEPLFGGKSCAOEMANPOHFL 1620  
1561 PIANGTVPGCPAKNVCDSNTCHNGGTCVNOMAPFSCBCEPLFGGKSCAOEMANPOHFL 1620  
1621 GSSLVAMHGLSLPISQPMYLSLMFRTROADGVLLQAITRGSTITTLQLRBGHVLVEGT 1680  
1621 GSSLVAMHGLSLPISQPMYLSLMFRTROADGVLLQAITRGSTITTLQLRBGHVLVEGT 1680  
1681 GLQASSRLRPRGRANDGDWHAQALGASGPGHAILSPYGOQRAEGNLGPRLHGLHS 1740  
1681 GLQASSRLRPRGRANDGDWHAQALGASGPGHAILSPYGOQRAEGNLGPRLHGLHS 1740  
1741 NITVGGIPGAGVYARFRGCLQGVRSPTPEGNSLDPSHGSINVEOGCSLDPDCSN 1800  
1741 NITVGGIPGAGVYARFRGCLQGVRSPTPEGNSLDPSHGSINVEOGCSLDPDCSN 1800  
1801 PCPANSYCSNDWMSYSCSDPGYGDNCCTNVCDLNPCEHOSVCTRKSPAHGYTCECPN 1860  
1801 PCPANSYCSNDWMSYSCSDPGYGDNCCTNVCDLNPCEHOSVCTRKSPAHGYTCECPN 1860  
1861 YLGPYCESTRIDQPCPRMWHPTCGPCNCVSKFDPDCKRTSGECHKENHYPPOSPT 1920  
1861 YLGPYCESTRIDQPCPRMWHPTCGPCNCVSKFDPDCKRTSGECHKENHYPPOSPT 1920  
1921 CLACDCYPTSLSRVCPEDGQCECKPGVIGROCDRCNPFABVTNICEVNYDSCRAI 1980  
1921 CLACDCYPTSLSRVCPEDGQCECKPGVIGROCDRCNPFABVTNICEVNYDSCRAI 1980  
1981 BAGIMWRTFRFGLPAAAPCEKGSFGTAVNRHCDERHGLPNNLFNCTSIITSEBLKFAERL 2040  
1981 BAGIMWRTFRFGLPAAAPCEKGSFGTAVNRHCDERHGLPNNLFNCTSIITSEBLKFAERL 2040  
2041 QRNESGLDSRSQLALLNNAOTHTAGYGSVYKVAVQATRLALRSTORRGLSATQ 2100  
2041 QRNESGLDSRSQLALLNNAOTHTAGYGSVYKVAVQATRLALRSTORRGLSATQ 2100  
2101 DVHFTENTLAVGSALDPTANKRMELIQTEGGTAMLLQHYEYASALQNMRTIYSPF 2160  
2101 DVHFTENTLAVGSALDPTANKRMELIQTEGGTAMLLQHYEYASALQNMRTIYSPF 2160  
2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGQPPLETTVILPESVFRETPPVVPAG 2220  
2161 TIVTPNIVISVRLDKGNFAGAKLPRYBALRGQPPLETTVILPESVFRETPPVVPAG 2220  
2221 PGEAQBEEELARORRHELSQGEAASVITTYTLGGLPHANDPDRSRIRVPRPIINT 2280  
2221 PGEAQBEEELARORRHELSQGEAASVITTYTLGGLPHANDPDRSRIRVPRPIINT 2280  
2281 PVSISVHDEBELLPALDQVTVQFRLLETERTKPICVFNMHSILVSTGMSARGCE 2340  
2281 PVSISVHDEBELLPALDQVTVQFRLLETERTKPICVFNMHSILVSTGMSARGCE 2340  
2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPKLTLYVALAGVTLALLTFPELTL 2400  
2341 VVFRNESHVSCQCNHMTSPAVLMDVSRRENGEILPKLTLYVALAGVTLALLTFPELTL 2400  
2401 LRIIRSNHGIRBNLRAALGLAQVFLIGNODLPACVIVAILLHFLVLCFPMALTE 2460

Db 2401 LRILRSNQHGRIRNRLTAALGLAQLVFLGLGINQADLPACTVIALILHFLYLCFSSMALLR 2460  
 Qy 2461 ALHLALYALTEVRDVTNTPMRFYMLGWSGPAFTGLAVGLDPGSGNPDGCMLSIVDTLLI 2520  
 Db 2461 ALHLALYALTEVRDVTNTPMRFYMLGWSGPAFTGLAVGLDPGSGNPDGCMLSIVDTLLI 2520  
 Qy 2521 WSPAGPVAFAVNSVFLYIYAARASCAAOQGFEGKGPVSGLOPSPAVILLISATWLLAL 2580  
 Db 2521 WSPAGPVAFAVNSVFLYIYAARASCAAOQGFEGKGPVSGLOPSPAVILLISATWLLAL 2580  
 Qy 2581 LSVNSDPLLFHYLPATCNCGIOPFPLSYVVLSEKVRKALKACSRKPSDPALTTKSTL 2640  
 Db 2581 LSVNSDPLLFHYLPATCNCGIOPFPLSYVVLSEKVRKALKACSRKPSDPALTTKSTL 2640  
 Qy 2641 TSSYNCPSPYADRLQPYDSDAGSLHSTSRSGKSPSYIPLLRBSALNPGQPGGLG 2700  
 Db 2641 TSSYNCPSPYADRLQPYDSDAGSLHSTSRSGKSPSYIPLLRBSALNPGQPGGLG 2700  
 Qy 2701 DPQSLPLEGDDQDHPDPTSDSLSLBDDQSGSYASTHSDSEEBEEREAAPFGEQG 2760  
 Db 2701 DPQSLPLEGDDQDHPDPTSDSLSLBDDQSGSYASTHSDSEEBEEREAAPFGEQG 2760  
 Qy 2761 WDSLPGGARLPLHSTPKDGGPGKAPWPGPTTAKSSGNGAPERRLRNGDALSR 2820  
 Db 2761 WDSLPGGARLPLHSTPKDGGPGKAPWPGPTTAKSSGNGAPERRLRNGDALSR 2820  
 Qy 2821 EGSGLPFGSSAOPHKGILKKKCLPTISEKSLRLPLBQCTGSSRGSSASBGSRGAPPP 2880  
 Db 2821 EGSGLPFGSSAOPHKGILKKKCLPTISEKSLRLPLBQCTGSSRGSSASBGSRGAPPP 2880  
 Qy 2881 RPPRRQSLQQLNGVWPIAMSIAGTYDESSGSEFLFFNPLH 2923  
 Db 2881 RPPRRQSLQQLNGVWPIAMSIAGTYDESSGSEFLFFNPLH 2923  
 RESULT 2  
 CLIR2\_MOUSE STANDARD; PRT; 2920 AA.  
 ID CLIR2\_MOUSE STANDARD; PRT; 2920 AA.  
 AC Q9ROM0; Q99K26; Q922R4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1) (Mfam11).  
 GN CELSR2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99418630; PubMed=10490098;  
 RA Uenai T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L., Takeichi M., Uemura T.;  
 RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of itz1/2.",  
 RL Cell 98:585-595(1999).  
 RN (2)  
 RP SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.  
 RX MEDLINE=20253755; PubMed=10790539;  
 RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;  
 RT "Chromosomal localization of Celser2 and Celser3 in the mouse; Celser3 is a candidate for the floppy (fip) lethal mutant on chromosome 9.",  
 RL Mamm. Genome 11:392-394(2000).  
 RN (3)  
 RP SEQUENCE OF 2014-2920 FROM N.A.  
 RC TISSUE-Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Strausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins S.F., Jordan H., Moore T., Max S.I., Wang J., Heien F., Datchenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carrincci P., Prange C., Raha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boase S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Meyers R.M.,  
 RA Butterfield Y.S.N., Krzywicki M.T., Skalska U., Smailus D.E., Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN (4)  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=21839555; PubMed=11850187;  
 RA Tissier F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;  
 RT "Developmental expression profiles of Celser (Flamingo) genes in the mouse.";  
 RL Mech. Dev. 112:157-160(2002).  
 CC -1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.  
 CC -1- DEVELOPMENTAL STAGE: Predominantly expressed in the developing CNS, the emerging dorsal root ganglia and cranial ganglia. In the CNS, expression is uniform along the rostrocaudal axis. During gastrulation, it is expressed within the anterior neural ectoderm. At E10, expression is strong in the ventricular zones (VZ) in all sectors of the brain, and lower in the marginal zones (MZ). Between E12 and E15, expression is prominent in the brain. It is strong in VZ, lower in MZ, except in telecephalic MZ where it is predominant. The intensity is higher in all VZ, and lower in differentiating fields than in VZ, except in the cerebral hemispheres, and to a lesser extent in the cecum and cerebellum. A weak expression is also observed in the fetal lungs, kidney and epithelia. In the newborn and postnatal stages, expression remains restricted to the VZ as well as in migrating and postmitogenic cells throughout the brain.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: Contains 9 cadherin domains.  
 CC -1- SIMILARITY: Contains 8 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin G-EGF-like domain.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AB028499; BAA84070.1; -  
 DR EMBL, AF031573; AAC68837.1; -  
 DR EMBL, BC005499; AAH05499.1; -  
 DR HSSP, P00740; 1EDM.  
 DR MGD, MGI:1858235; Celser2.  
 DR GO, GO:0016021; C:integral to membrane, ISS.  
 DR GO, GO:0004930; F:G-protein coupled receptor activity, ISS.  
 DR InterPro, IPR000152; Asx\_hydroxyl.  
 DR InterPro, IPR002126; Cadherin.  
 DR InterPro, IPR000742; EGF 2.  
 DR InterPro, IPR006209; EGF-like.  
 DR InterPro, IPR000832; GPCR\_secretin.  
 DR InterPro, IPR001879; hormn\_receptor.  
 DR InterPro, IPR002049; laminin\_EGF.  
 DR InterPro, IPR001791; laminin\_G.  
 DR InterPro, IPR000203; PKD\_cys\_rich.  
 DR Pfam, PF00002; 7cm\_2; 1.

DR Pfam; PF000028; cadherin; 9.  
 DR Pfam; PF000008; EGF; 5.  
 DR Pfam; PF01825; GBS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00053; laminin\_EGF; 1.  
 DR Pfam; PF00054; laminin\_G; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00180; EGF\_Lam; 1.  
 DR SMART; SM00303; GBS; 1.  
 DR SMART; SM00008; Horrn; 1.  
 DR SMART; SM00282; Lamg; 2.  
 DR PROSITE; PS00010; ASK\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 6.  
 DR PROSITE; PS00268; CADHERIN\_2; 9.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR PROSITE; PS00221; GBS; 1.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;  
 DR EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 DR Developmental protein; Hydroxylation; Signal.  
 DR SIGNAL 1 31  
 FT CHAIN 32 2920  
 FT  
 FT DOMAIN 32 2381  
 FT TRANSMEM 2382 2402  
 FT TRANSMEM 2403 2414  
 FT TRANSMEM 2415 2434  
 FT TRANSMEM 2435 2439  
 FT TRANSMEM 2440 2460  
 FT TRANSMEM 2461 2481  
 FT TRANSMEM 2482 2502  
 FT TRANSMEM 2503 2519  
 FT TRANSMEM 2520 2540  
 FT TRANSMEM 2541 2564  
 FT TRANSMEM 2565 2585  
 FT TRANSMEM 2586 2592  
 FT TRANSMEM 2593 2613  
 FT TRANSMEM 2614 2920  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 506  
 FT DOMAIN 507 611  
 FT DOMAIN 612 713  
 FT DOMAIN 714 816  
 FT DOMAIN 817 922  
 FT DOMAIN 923 1024  
 FT DOMAIN 1029 1147  
 FT DOMAIN 1229 1287  
 FT DOMAIN 1289 1319  
 FT DOMAIN 1329 1367  
 FT DOMAIN 1368 1572  
 FT DOMAIN 1575 1611  
 FT DOMAIN 1615 1792  
 FT DOMAIN 1788 1830  
 FT DOMAIN 1831 1923  
 FT DOMAIN 1884 1933  
 FT DOMAIN 1924 1956  
 FT DOMAIN 2317 2369  
 FT DOMAIN 2744 2749  
 FT DISULFID 1293 1304  
 FT DISULFID 1298 1313  
 FT DISULFID 1315 1318  
 FT DISULFID 1333 1344  
 FT DISULFID 1338 1354

FT DISULFID 1356 1366 BY SIMILARITY.  
 FT DISULFID 1579 1590 BY SIMILARITY.  
 FT DISULFID 1584 1599 BY SIMILARITY.  
 FT DISULFID 1601 1610 BY SIMILARITY.  
 FT DISULFID 1792 1803 BY SIMILARITY.  
 FT DISULFID 1798 1818 BY SIMILARITY.  
 FT DISULFID 1820 1829 BY SIMILARITY.  
 FT DISULFID 1833 1844 BY SIMILARITY.  
 FT DISULFID 1838 1856 BY SIMILARITY.  
 FT DISULFID 1858 1867 BY SIMILARITY.  
 FT DISULFID 1888 1900 BY SIMILARITY.  
 FT DISULFID 1890 1907 BY SIMILARITY.  
 FT DISULFID 1909 1922 BY SIMILARITY.  
 FT DISULFID 1925 1937 BY SIMILARITY.  
 FT DISULFID 1927 1944 BY SIMILARITY.  
 FT DISULFID 1946 1955 BY SIMILARITY.  
 FT MOD RES 1592 1592 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 486 486 N-LINKED (GLCNAC...) (POTENTIAL).  
 Query Match 94.3%; Score 14657.5; DB 1; length 2920;  
 Best Local Similarity 94.4%; Pred. No. 0;  
 Matches 2760; Conservative 54; Mismatches 105; Indels 5; Gaps 2;  
 QY 1 MRSATGVPLPPTPPPLLLLLLLLLPPPLGDDVGPCHSLGSRGSGGACAPMGWLCPS 60  
 DB 1 MRTAASAPLPPTPLPLPLLLLLLPPPLGDDVGPCHSLGSGRSGSGACAPMGWLCPS 60  
 QY 61 SASNIMLYTSRCRAGTELNGHLVPHDGLRWCPESBAHPLPAPGCPMGCRLGIG 120  
 DB 61 SASNIMLYTSRCRSGGIELTHLVPHDGLRWCPESBAHPLPAPSGGCPMGCRLGIG 120  
 QY 121 GHLSPOGKTLPEHPCLKAPRLCQSCCKLAQAPGLRAGERSPEBSLGRKRVNTAPQ 180  
 DB 121 GHLSPOGKTLPEHPCLKAPRLCQSCCKLAQAPGLRAGERSPEBSLGRKRVNTAPQ 180  
 QY 181 POPSYGATVBNOPACTPPVASLAIIDPDSGARLRYTMDALDSSNNOFSLDPVTGA 240  
 DB 181 POPSYGATVBNOPACTPPVASLAIIDPDSGARLRYTMDALDSSNNOFSLDPVTGA 240  
 QY 241 VTTEELDRKTSFHFVRVTAQDHGMPRSALATLTTLVTDTNHDVPEGOBYKESLRE 300  
 DB 241 VTTEELDRKTSFHFVRVTAQDHGMPRSALATLTTLVTDTNHDVPEGOBYKESLRE 300  
 QY 301 NLEVGKTLVRAADGAPPPANILYRLBGSQSPSEVEIDRSGVIRTRGVDRREV 360  
 DB 301 NLEVGKTLVRAADGAPPPANILYRLBGSQSPSEVEIDRSGVIRTRGVDRREV 360  
 QY 361 ESYGLTVEASDQGRDPPRSTALVFLSVEDDNDNAQPFSEKRVYVVRBDVTGAPVLR 420  
 DB 361 ESYGLTVEASDQGRDPPRSTALVFLSVEDDNDNAQPFSEKRVYVVRBDVTGAPVLR 420  
 QY 421 VTASDRDKSGNVAHVHISGNGARQFYLDAGTALVVSPLDYETTKYTLVRAQDGG 480  
 DB 421 VTASDRDKSGNVAHVHISGNGARQFYLDAGTALVVSPLDYETTKYTLVRAQDGG 480  
 QY 481 RPLPSNVSGLVTVQVLDIND-NAPIFVSTPPQATVLSVPLGYLVAVQALDADAGNAR 539  
 DB 481 RPLPSNVSGLVTVQVLDIND-NAPIFVSTPPQATVLSVPLGYLVAVQALDADAGNAR 539  
 QY 540 LEYLLAAGVDPPTTNGGWTISVAELDRREVDFPSFGVBAADHGTPLTASASVYT 599  
 DB 540 LEYLLAAGVDPPTTNGGWTISVAELDRREVDFPSFGVBAADHGTPLTASASVYT 599  
 QY 541 LEYLLAAGVDPPTTNGGWTISVAELDRREVDFPSFGVBAADHGTPLTASASVYT 600  
 DB 541 LEYLLAAGVDPPTTNGGWTISVAELDRREVDFPSFGVBAADHGTPLTASASVYT 600  
 QY 600 VLDVNDNNPTFTQDEYTVRLNEDAAVGTSVTVSAVDRDAHSVITYOITSNTNRFSIT 659  
 DB 601 ILVDNDNNPTFTQDEYTVRLNEDAAVGTSVTVSAVDRDAHSVITYOITSNTNRFSIT 660  
 QY 660 SOSGGGLVSLALPLDYTLERQYLAVTASDGTQDTAQLIVNVVTDANTHRPVFOSSHYTV 719  
 DB 661 SOSGGGLVSLALPLDYTLERQYLAVTASDGTQDTAQLIVNVVTDANTHRPVFOSSHYTV 720  
 QY 720 NVNEDRAGTTVVLIATSDGTGEMARITTFMBSIIOFPLIDDTGAVTQAGLIDYDDOV 779  
 DB 720 NVNEDRAGTTVVLIATSDGTGEMARITTFMBSIIOFPLIDDTGAVTQAGLIDYDDOV 779

Db 721 NGNEBPAQTVVLLISATDEDTGENARITYFMEBDSIPQFRIDGDTGAVTTQAEIYEDQV 780  
 Qy 780 SYTLATIPARNGI POKSDTTYLEILVNDVMDNAQFRLDSYGOSVYEDVPPTSYQISA 839  
 Db 781 SYTLATIPARNGI POKSDTTYLEILVNDVMDNAQFRLDSYGOSVYEDVPPTSYQISA 840  
 Qy 840 TDRDSGLNGEVFTYFQGGDDGDDGFIVESTGIVRTLRRLRENVAAQYVLAAYDKMP 899  
 Db 841 TDRDSGLNGEVFTYFQGGDDGDDGFIVESTGIVRTLRRLRENVAAQYVLAAYDKMP 900  
 Qy 900 PARTPEVTVTVLDVNDNPNVFEDEDFVVEENSPIGLAVARTATADPEGTNAQIMYO 959  
 Db 901 PARTPEVTVTVLDVNDNPNVFEDEDFVVEENSPIGLAVARTATADPEGTNAQIMYO 960  
 Qy 960 IVEGNIPEVFOQDFISGELTALVDLDYEDRPEYVLTQATSAPVSRATVHRLDNDN 1019  
 Db 961 IVEGNIPEVFOQDFISGELTALVDLDYEDRPEYVLTQATSAPVSRATVHRLDNDN 1020  
 Qy 1020 PVLGNFEILFNNVYTRSSSPGGAIGRVPAHDPIISDSLTVSFERGNEILSVLNAST 1079  
 Db 1021 PVLGNFEILFNNVYTRSSSPGGAIGRVPAHDPIISDSLTVSFERGNEILSVLNAST 1080  
 Qy 1080 GELKLSALNNPRLAIVMSVSDGYHSTAQCALAVTTITDBMLTHSTITLLEDMSPR 1139  
 Db 1081 GELKLSALNNPRLAIVMSVSDGYHSTAQCALAVTTITDBMLTHSTITLLEDMSPR 1140  
 Qy 1140 RFLSPLGLFIOAVATLTPPDHVVFNQRTDAPGHIILNVLSVQPPGPGGPP 1199  
 Db 1141 RFLSPLGLFIOAVATLTPPDHVVFNQRTDAPGHIILNVLSVQPPGPGGPP 1200  
 Qy 1200 LPEBDLOERLYLRSLITLTAISAQVLPDPDNICLREPCENYMCVSVLRFSSAPFIASS 1259  
 Db 1201 LPEBDLOERLYLRSLITLTAISAQVLPDPDNICLREPCENYMCVSVLRFSSAPFIASS 1260  
 Qy 1260 SVLFRPHVPGARCRCPGPTGDTCEYDLCVSRCPGHRCSRSEGGYTCLCEDGYT 1319  
 Db 1261 SVLFRPHVPGARCRCPGPTGDTCEYDLCVSRCPGHRCSRSEGGYTCLCEDGYT 1320  
 Qy 1320 GEHCENVSARSGRCTPGVCKNGGTCVNLVAGAFKDCPSGDPBEKPYCVATTTRSPAFHSFT 1379  
 Db 1321 GEHCENVSARSGRCTPGVCKNGGTCVNLVAGAFKDCPSGDPBEKPYCVATTTRSPAFHSFT 1380  
 Qy 1380 FRGLRQRFHTLALSPATKRDGILLNRPNRKHPVALLEVIOEVOULTFSAGESTTV 1439  
 Db 1381 FRGLRQRFHTLALSPATKRDGILLNRPNRKHPVALLEVIOEVOULTFSAGESTTV 1440  
 Qy 1440 SPFVPGSVSGOMHTVOLKTYNKPILGQGTGLPOGSEBQKXAVTVTVBSCDGVALLRGSVY 1499  
 Db 1441 SPFVPGSVSGOMHTVOLKTYNKPILGQGTGLPOGSEBQKXAVTVTVBSCDGVALLRGSVY 1500  
 Qy 1500 GNTSCAAQGTQGGSKSLDTGLPILGQVPLDPESFVRMRQFVGCWRNLQVDSRHIDMA 1559  
 Db 1501 GNTSCAAQGTQGGSKSLDTGLPILGQVPLDPESFVRMRQFVGCWRNLQVDSRHIDMA 1560  
 Qy 1560 DFLANNGTVPGCPAKKXVCSNCTCANGGTCVNDAPSCCEPIJFGGKSQAQEMANPOH 1619  
 Db 1561 DFLANNGTVPGCPAKKXVCSNCTCANGGTCVNDAPSCCEPIJFGGKSQAQEMANPOH 1620  
 Qy 1620 LGSGLVAMHGLSLPISQPTWILMFTTRQADGVTLQATIRGRTITTLQLEBGMVLSVEG 1679  
 Db 1621 LGSGLVAMHGLSLPISQPTWILMFTTRQADGVTLQATIRGRTITTLQLEBGMVLSVEG 1680  
 Qy 1680 TGLQASSLRLPERGANDGDWHHAQALGASGPGHAILSPDYQQAAGNLAGRLHGLHL 1739  
 Db 1681 TGLQASSLRLPERGANDGDWHHAQALGASGPGHAILSPDYQQAAGNLAGRLHGLHL 1740  
 Qy 1740 SNTTVGSLRBPAGVAVAGFRGCLQGVVSDTPBGSVNLDSHGBSINVEGCSLLPDCDS 1799  
 Db 1741 SNTTVGSLRBPAGVAVAGFRGCLQGVVSDTPBGSVNLDSHGBSINVEGCSLLPDCDS 1800  
 Qy 1800 NPCRANSYCSNDMDYSCSGCPGYGDNCTNVCDLANCEHQSVCTRKPSAPHGYCECP 1859  
 Db 1801 NPCRANSYCSNDMDYSCSGCPGYGDNCTNVCDLANCEHQSVCTRKPSAPHGYCECP 1860

Qy 1860 NYLGPYCETRIDOPCRGMWGHPTGCPNCNDVSKGFPDPCNKTSGECHKENHYRPPGSP 1919  
 Db 1861 NYLGPYCETRIDOPCRGMWGHPTGCPNCNDVSKGFPDPCNKTSGECHKENHYRPPGSP 1920  
 Qy 1920 TGLCDYPTGSLSRVCDDEGQCPCKPVIGROCDRCNPPAVTTNCEVNYDSCPA 1979  
 Db 1921 TGLCDYPTGSLSRVCDDEGQCPCKPVIGROCDRCNPPAVTTNCEVNYDSCPA 1980  
 Qy 1980 IEAGIWPBTRPGALPAAPCPGSGTAVRCHDERGMLPMLFNCTSTSEBELGFAAR 2039  
 Db 1981 IEAGIWPBTRPGALPAAPCPGSGTAVRCHDERGMLPMLFNCTSTSEBELGFAAR 2040  
 Qy 2040 LQRNESGLDSGRSQQALALLRRAOTHTAGYFGSDVKAVALQTLATRLLAHSTORFGLSAT 2099  
 Db 2041 LQRNESGLDSGRSQQALALLRRAOTHTAGYFGSDVKAVALQTLATRLLAHSTORFGLSAT 2100  
 Qy 2100 QVHFTENILRVGSALLDANRGMELIQTEGRTAMLLQHYEAYASALAQNMRHTYLS 2159  
 Db 2101 QVHFTENILRVGSALLDANRGMELIQTEGRTAMLLQHYEAYASALAQNMRHTYLS 2160  
 Qy 2160 FTIVTPIVIVSVRLDKGNFAGAKLPRYALRGOBPDLETTVILPESVFRBTPPVVRA 2219  
 Db 2161 FTIVTPIVIVSVRLDKGNFAGAKLPRYALRGOBPDLETTVILPESVFRBTPPVVRA 2220  
 Qy 2220 GPGEAQEPBELARORRHPBELSGEAVASVIIYRTLAGLPHNYDPDKRSLRVPKPIIN 2279  
 Db 2221 GPGEAQEPBELARORRHPBELSGEAVASVIIYRTLAGLPHNYDPDKRSLRVPKPIIN 2280  
 Qy 2280 TVVVISVHDEBELPRALDKVTVYQFRLLEBETKPICVWNHSLVSGTGMASRG 2339  
 Db 2281 TVVVISVHDEBELPRALDKVTVYQFRLLEBETKPICVWNHSLVSGTGMASRG 2340  
 Qy 2340 EYVFRNESHVSQCNHMTSFAVLMDSVRBENGELPLKTLVYVAGVTLAALLTFEFT 2399  
 Db 2341 EYVFRNESHVSQCNHMTSFAVLMDSVRBENGELPLKTLVYVAGVTLAALLTFEFT 2400  
 Qy 2400 LRLIRSNQHGIRNLTALGLAOLVFLGINOQDLPACTVIAITLHFLYCTFSWALL 2459  
 Db 2401 LRLIRSNQHGIRNLTALGLAOLVFLGINOQDLPACTVIAITLHFLYCTFSWALL 2460  
 Qy 2460 EALHLYRALTETVDNTPGMRPYVNLGCVAPFITGLAVGLDPRGYNDDPCWLSYDTL 2519  
 Db 2461 EALHLYRALTETVDNTPGMRPYVNLGCVAPFITGLAVGLDPRGYNDDPCWLSYDTL 2520  
 Qy 2520 IWSFAGPVAFVMSVFLYITLAAASCAAQOGFEKKGVSGLQSPFAVTLILSATWILIA 2579  
 Db 2521 IWSFAGPVAFVMSVFLYITLAAASCAAQOGFEKKGVSGLQSPFAVTLILSATWILIA 2580  
 Qy 2580 LLSVNSDTLLFHYLPATNCIQGPPIFLSYVVLSEVRKALKIACSRKSPDPALTTKST 2639  
 Db 2581 LLSVNSDTLLFHYLPATNCIQGPPIFLSYVVLSEVRKALKIACSRKSPDPALTTKST 2640  
 Qy 2640 LTRSYNCSBPYADGSLYOPYGSAGSLHSTRSGSOSBYIFPLREBSALNPGQGPCL 2699  
 Db 2641 LTRSYNCSBPYADGSLYOPYGSAGSLHSTRSGSOSBYIFPLREBSALNPGQGPCL 2700  
 Qy 2700 GDPGSLFLEGQDQDPTDSDSLLEDDGSGSASTHSDSEBEEBEEBAAFPGEQ 2759  
 Db 2701 GDPGSLFLEGQDQDPTDSDSLLEDDGSGSASTHSDSEBEEBEEBAAFPGEQ 2760  
 Qy 2760 GWDLSLIGPAERLPLHSTPKDGGPGPKAPMGFDGTAKESSGNGAPREBLRENGDLS 2819  
 Db 2761 GWDLSLIGPAERLPLHSTPKDGGPGPKAPMGFDGTAKESSGNGAPREBLRENGDLS 2820  
 Qy 2820 REGSLGPIBPGSSAOHKKGLKKCLPTLSKXSLRLPLBOCTGSSRSGSASBEGSGGPP 2879  
 Db 2821 REGSLGPIBPGSSAOHKKGLKKCLPTLSKXSLRLPLBOCTGSSRSGSASBEGSGGPP 2880  
 Qy 2880 PRPPRQSLQBOQNGVMPVAMSIRKAGTVDBDSGSGEPLFFNPLH 2923  
 Db 2877 PRPPRQSLQBOQNGVMPVAMSIRKAGTVDBDSGSGEPLFFNPLH 2920





```

FT CARBOHYD 261 261 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 726 726 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 966 966 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1249 1249 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1268 1268 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1286 1286 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1548 1548 N-LINKED (GLCNAC. .) (POTENTIAL. .)
FT CARBOHYD 1570 1570 N-LINKED (GLCNAC. .) (POTENTIAL. .)
SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;

Query Match 71.4%; Score 11097; DB 1; Length 2144;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 2062; Conservative 35; Mismatches 47; Indels 4; Gaps 1;

Qy 776 EDVSYTLATARDNGIPKSDTYLEILVNDVNDNAPOFLRDSYQGSVEEDVPFTSVL 835
Db 1 EDVSYTLATARDNGIPKSDTYLEILVNDVNDNAPOFLRDSYQGSVEEDVPFTSVL 60
Qy 836 QISATRDSGLNGRVFTYFQGGDGDGFVSTSGIVRTLRDLRENAQYVLRAYAD 895
Db QISATRDSGLNGRVFTYFQGGDGDGFVSTSGIVRTLRDLRENAQYVLRAYAD 120
Qy 896 KGNPAPATPMEVYTVYLDVNDNPFVEFODEFVDVENSPIGLAVAVATTDDEGNAQ 955
Db KGNPAPATPMEVYTVYLDVNDNPFVEFODEFVDVENSPIGLAVAVATTDDEGNAQ 180
Qy 956 IMYQIVENGIPEVFOLDIPFSGELTALVDLYEDREPEVLYLOATSAPLVRATVHVL 1015
Db IMYQIVENGIPEVFOLDIPFSGELTALVDLYEDREPEVLYLOATSAPLVRATVHVL 240
Qy 1016 RNDNPPVLGNPELIFNNYVTRNSSSPFGAIGRVPAHDPDISGLTVSPFERGELSLV 1075
Db RNDNPPVLGNPELIFNNYVTRNSSSPFGAIGRVPAHDPDISGLTVSPFERGELSLV 300
Qy 1076 NASTGELKLRALDNNPLRAIMSVLSDGVHSTACALRVTTIDEMTHSTLTLED 1135
Db NASTGELKLRALDNNPLRAIMSVLSDGVHSTACALRVTTIDEMTHSTLTLED 360
Qy 1136 MSPEPLPLPLGLFIQAVATLATPDPHVVFNVOQRTDAPGGHILNVSLVQCPFGG 1195
Db MSPEPLPLPLGLFIQAVATLATPDPHVVFNVOQRTDAPGGHILNVSLVQCPFGG 420
Qy 1196 GPPPLPEBDLQERLYLNRSLTALSAQVLPEDDNICLRBPCENYMCVSLRPSGAP 1255
Db GPPPLPEBDLQERLYLNRSLTALSAQVLPEDDNICLRBPCENYMCVSLRPSGAP 480
Qy 1256 IASSSVFRPLHPVGLRCRCPGFTGDCETEDVLCYRPGPHGCRSRBGTYLCL 1315
Db IASSSVFRPLHPVGLRCRCPGFTGDCETEDVLCYRPGPHGCRSRBGTYLCL 540
Qy 1316 DGYTGEHCEVSARSRCCTPGVCKNGTCVNLVGFKCDPSPGDFEKPQCVTTSPFAH 1375
Db DGYTGEHCEVSARSRCCTPGVCKNGTCVNLVGFKCDPSPGDFEKPQCVTTSPFAH 600
Qy 1376 SFTIFRGLRQRFHTTALSPATKRDGLLYNGRFNKHDFVALLEVIOEOVLFTSAGES 1435
Db SFTIFRGLRQRFHTTALSPATKRDGLLYNGRFNKHDFVALLEVIOEOVLFTSAGES 660
Qy 1436 TTYTSPFVPGGVSGQWHTVQKTKYNKPLGQGTLPQSPSOKAVAVTVGCDTVALRF 1495
Db TTYTSPFVPGGVSGQWHTVQKTKYNKPLGQGTLPQSPSOKAVAVTVGCDTVALRF 720
Qy 1496 GSVLGNYSCAAQGTQGSKSLDTJGPLLGGVLDLPESFPVRNRQFVGCNRLQVDSRH 1555
Db GSVLGNYSCAAQGTQGSKSLDTJGPLLGGVLDLPESFPVRNRQFVGCNRLQVDSRH 780
Qy 1556 IDMAFDPIANNCTVGCRAKKNVCDSNTCHNGTCVNMWDAPSCGCPGLGPGKSCAQEMAN 1615

```

```

Db 781 VDMAFDPIANNCTVGCRAKKNVCDSNTCHNGTCVNMWDAPSCGCPGLGPGKSCAQEMAN 840
Qy 1616 PQHFLASSLVANHGSLPISQPMYLSMERTQADGVLAQITTRGSTITTLQREHVL 1675
Db PQHFLASSLVANHGSLPISQPMYLSMERTQADGVLAQITTRGSTITTLQREHVL 900
Qy 1676 SVEGTGLQASSLRLEGRANDGMHHAQALASGGPGHAIISFDVGOORAEINLGRHL 1735
Db SVEGTGLQASSLRLEGRANDGMHHAQALASGGPGHAIISFDVGOORAEINLGRHL 960
Qy 1736 GLHLSNTVGIIPGAPGAVARFGTQGVRSVDPFEGVNSLDPHSGSINVEQCSLPD 1795
Db GLHLSNTVGIIPGAPGAVARFGTQGVRSVDPFEGVNSLDPHSGSINVEQCSLPD 1020
Qy 1796 PCDSNPCPANSYCSNDMDVSYSCSDPGYVGDCTVNCIDLPCEHOSACTRKSAPRGYTC 1855
Db PCDSNPCPANSYCSNDMDVSYSCSDPGYVGDCTVNCIDLPCEHOSACTRKSAPRGYTC 1080
Qy 1856 ECPNVLGPGYCETRLDPQCFRGMWGHPTGPNCDVSKGFPDPCNKTSGECHKENHYP 1915
Db ECPNVLGPGYCETRLDPQCFRGMWGHPTGPNCDVSKGFPDPCNKTSGECHKENHYP 1140
Qy 1916 PESPFTCLDCYPTGSLRVCDPBDGQCEKRGVIGRQCDRCNDPFAVYTTNGCEVNYDS 1975
Db PESPFTCLDCYPTGSLRVCDPBDGQCEKRGVIGRQCDRCNDPFAVYTTNGCEVNYDS 1200
Qy 1976 CPRAIENAGIWMPTRTGFLPAAAPCPKSGRGTAVRCHDERGMLPPLFCTSTITSBELG 2035
Db CPRAIENAGIWMPTRTGFLPAAAPCPKSGRGTAVRCHDERGMLPPLFCTSTITSBELG 1260
Qy 2036 FAERLQRNESGLDSRQOLALLRNATQHTAGVSGDVKAVYQALATRLAHESTORFG 2095
Db FAERLQRNESGLDSRQOLALLRNATQHTAGVSGDVKAVYQALATRLAHESTORFG 1320
Qy 2096 LSATQDVHTENULRAYGALLDTPANKRMELIQTBEGTAMLLQHYEAYASALAQMRT 2155
Db LSATQDVHTENULRAYGALLDTPANKRMELIQTBEGTAMLLQHYEAYASALAQMRT 1380
Qy 2156 YLSPFTVFNVISVVRDLKGNPAGAKLPRYBALRGEPDPLETTVILPESVFRTEPV 2215
Db YLSPFTVFNVISVVRDLKGNPAGAKLPRYBALRGEPDPLETTVILPESVFRTEPV 1440
Qy 2216 VBPAGGEAQBEBELARRQRRPBELOSBAVASIITRTLAGLPHNYDPDRSLRVPKR 2275
Db VBPAGGEAQBEBELARRQRRPBELOSBAVASIITRTLAGLPHNYDPDRSLRVPKR 1500
Qy 2276 PIIINTPVVISVHDBBELPRALDYPVYQFRLTTEERTKEICVFMNHSILVSGTGM 2335
Db PIIINTPVVISVHDBBELPRALDYPVYQFRLTTEERTKEICVFMNHSILVSGTGM 1560
Qy 2336 ARGCEVFNRESHVSQCNHMTSPAVLMDVSRRENGEILPLKTLFTVALGVTLLALTF 2395
Db ARGCEVFNRESHVSQCNHMTSPAVLMDVSRRENGEILPLKTLFTVALGVTLLALTF 1620
Qy 2396 FPLTLRLIRNSNOHGRIRNLTALGLAQVLLGINQADLPACTVIALTLHFLYCTES 2455
Db FPLTLRLIRNSNOHGRIRNLTALGLAQVLLGINQADLPACTVIALTLHFLYCTES 1680
Qy 2456 WALLEMLHLYRALTERVDVNTGPNRYTNGVGAFTIGLAVGDDPEYGNPDCMLSI 2515
Db WALLEMLHLYRALTERVDVNTGPNRYTNGVGAFTIGLAVGDDPEYGNPDCMLSI 1740
Qy 2516 YDTLINSFAGPVAFAVMSVFLYIILABASCAQORGFEEKGPVSGLQSPFVILLLSAT 2575
Db YDTLINSFAGPVAFAVMSVFLYIILABASCAQORGFEEKGPVSGLQSPFVILLLSAT 1800
Qy 2576 WLIALLSVNSDTLLFHTYLFATNCIOGPFIPLSYVVLSEVRKALKALCSRPSPDALT 2635
Db WLIALLSVNSDTLLFHTYLFATNCIOGPFIPLSYVVLSEVRKALKALCSRPSPDALT 1860
Qy 2636 TKSITLSSVNCSPVADGRLVQPYGDSAGSLSTSRSGSOPSYIPFLIRRESALNPGOG 2695

```

```

Db 1861 TKTSTLSSVNCSPYADGRXYGDSAGSLHNSHSGSQPYTLPILRBSSTLNGOV 1920
Qy 2696 PPGGLDPSGLFLEGQDQDHPDSDSLSDLDQSGSVASTHSDSEEEEEEEAAAF 2755
Db 1921 PPGGLDPSGLFMEGQAQDHPDSDSLSDLDQSGSVASTHSDS---EEEEAAAF 1976
Qy 2756 PPGQGDSDLLGPAEAEPLPLHSTKDGPGGKAPWQDPTTAKESGNGCAPERLRENG 2815
Db 1977 PPGQGDSDLLGPAEAEPLPLHSTKDGPGGKAPWQDPTTAKESGNGCAPERLRENG 2036
Qy 2816 DALSRSGSLGPLPGSSAQPHKGLKKKCLPTISEKSLTLRLPLRQCGSSRGSASRGR 2875
Db 2037 DALTRSGSLGPLPGSTQPHKGLKKKCLPTISEKSLTLRLPLRQCGSSRGSASRGR 2096
Qy 2876 GGGPPPPPPRQSLQEQLVNVPJMSIKACTVDEDSGSEFLPFNTLH 2923
Db 2097 NGPPPPPPRQSLQEQLVNVPJMSIKACTVDEDSGSEFLPFNTLH 2144

RESULT 4
CLRI MOUSE STANDARD; PRT; 3034 AA.
AC 035161;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caderin EGF IAG seven-pass G-type receptor 1 precursor.
GN CELSR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9077689; PubMed=9858697;
RA Hadjantonakis A.-K., Formstone C.J., Little P.F.R.;
RT "Celsr1 is an evolutionarily conserved seven-pass transmembrane
RT receptor and is expressed during mouse embryonic development.";
RL Mech. Dev. 78:91-95(1998).
RN (2)
RP TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=97480720; PubMed=9339365;
RA Hadjantonakis A.-K., Sheward W.J., Hartmar A.J., de Galan L.,
RA Hoovers J.M.N., Little P.F.R.;
RT "Celsr1, a neural-specific gene encoding an unusual seven-pass
RT transmembrane receptor, maps to mouse chromosome 15 and human
RT chromosome 22qter.";
RL Genomics 45:97-104(1997).
RN (3)
RP DEVELOPMENTAL STAGE.
RX MEDLINE=21839555; PubMed=11850187;
RA Tissier P., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
RT "Developmental expression profiles of Celer (Flamingo) genes in the
RT mouse.";
RL Mech. Dev. 112:157-160(2002).
RN (4)
RP FUNCTION: Receptor that may have an important role in cell/cell
RP signaling during nervous system formation.
RN (5)
RP SUBCELLULAR LOCATION: Integral membrane protein.
RN (6)
RP TISSUE SPECIFICITY: Expressed in the brain, where it is localized
RP principally in the ependymal cell layer, choroid plexus and the
RP area postrema. Also found in spinal cord and in the eye.
RN (7)
RP DEVELOPMENTAL STAGE: First detected at E6. Predominantly expressed
RP in the developing CNS, the emerging dorsal root ganglia and
RP cranial ganglia. In the CNS, expression is uniform along the
RP rostrocaudal axis. During gastrulation, it is expressed in the
RP vicinity of the primitive streak, and becomes predominant in the
RP area at late gastrulation. At E10, detected in ventricular zones
RP (VZ), but not in marginal zones (MZ), and weakly in other
RP structures. Between E12 and E15, a high expression is present in
RP the VZ in all brain areas. No expression in differentiated
RP neuronal fields. In the newborn and postnatal stages, expression
RP remains restricted to the VZ. Also found weakly in fetal lungs,

```

```

CC kidney and epithelia.
CC -1 SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1 SIMILARITY: Contains 9 caderin domains.
CC -1 SIMILARITY: Contains 8 EGF-like domains.
CC -1 SIMILARITY: Contains 2 laminin G-like domains.
CC -1 SIMILARITY: Contains 1 laminin G-like domain.
CC -1 SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1 SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF031572; AAC68836.1; -.
DR PIR, T14119; T14119.
DR MGD, MGI:1106883; Celsr1.
DR HSSP, P00749; IURK.
DR InterPro, IPR000152; Asx hydroxyl.
DR InterPro, IPR002126; Caderin.
DR InterPro, IPR006209; EGF like.
DR InterPro, IPR000832; GPCR_secretin.
DR InterPro, IPR001879; horum_receptor.
DR InterPro, IPR002049; laminin_G.
DR InterPro, IPR001791; laminin_G.
DR InterPro, IPR000203; PKD_cys_rich.
DR pfam, PF000002; 7tm 2. 1.
DR pfam, PF000028; caderin. 9.
DR pfam, PF00008; EGF. 6.
DR pfam, PF01825; GPS. 1.
DR pfam, PF02793; HRM. 1.
DR pfam, PF00053; laminin_EGF. 1.
DR pfam, PF00054; laminin_G. 1.
DR PRINTS, PR00205; CADERIN.
DR PRINTS, PR00011; EGF_LAMININ.
DR PRINTS, PR00249; GPCRSECRETIN.
DR SMART, SM00112; CA. 9.
DR SMART, SM00180; EGF_Lam. 1.
DR SMART, SM00303; GPS. 1.
DR SMART, SM00008; Hornm. 1.
DR SMART, SM00282; LamG. 2.
DR PROSITE, PS00010; ASX_HYDROXYL. 2.
DR PROSITE, PS00232; CADERIN_1; 7.
DR PROSITE, PS00268; CADERIN_2; 9.
DR PROSITE, PS00022; EGF_1; 6.
DR PROSITE, PS01186; EGF_2; 2.
DR PROSITE, PS50221; GPS. 1.
DR PROSITE, PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE, PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE, PS50227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE, PS50261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE, PS01248; LAMININ_TYPE_EGF. 1.
DR PROSITE, PS50025; Lam_G_DOMAIN; 2.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 29
FT CHAIN 30 3034
FT DOMAIN 21 2484
FT TRANSMEM 2485 2505
FT DOMAIN 2506 2516
FT TRANSMEM 2517 2537
FT DOMAIN 2538 2542
FT TRANSMEM 2543 2563
FT DOMAIN 2564 2587
FT TRANSMEM 2588 2608
FT DOMAIN 2609 2625
FT TRANSMEM 2626 2646
FT DOMAIN 2647 2670
FT TRANSMEM 2671 2691

```

FT DOMAIN 2692 2694 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 2695 2715 7 (POTENTIAL).  
 FT DOMAIN 2716 3034 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 261 368 CADHERIN 1.  
 FT DOMAIN 369 474 CADHERIN 2.  
 FT DOMAIN 475 580 CADHERIN 3.  
 FT DOMAIN 581 702 CADHERIN 4.  
 FT DOMAIN 703 804 CADHERIN 5.  
 FT DOMAIN 805 907 CADHERIN 6.  
 FT DOMAIN 908 1014 CADHERIN 7.  
 FT DOMAIN 1015 1116 CADHERIN 8.  
 FT DOMAIN 1121 1239 CADHERIN 9.  
 FT DOMAIN 1318 1376 EGF-LIKE 1, CALCIUM-BINDING.  
 FT DOMAIN 1378 1414 EGF-LIKE 2, CALCIUM-BINDING.  
 FT DOMAIN 1418 1456 EGF-LIKE 3, CALCIUM-BINDING.  
 FT DOMAIN 1457 1661 LAMININ G-LIKE 1.  
 FT DOMAIN 1664 1700 EGF-LIKE 4, CALCIUM-BINDING.  
 FT DOMAIN 1704 1885 LAMININ G-LIKE 2.  
 FT DOMAIN 1887 1922 EGF-LIKE 5, CALCIUM-BINDING.  
 FT DOMAIN 1923 1961 EGF-LIKE 6, CALCIUM-BINDING.  
 FT DOMAIN 1962 1994 EGF-LIKE 7, CALCIUM-BINDING.  
 FT DOMAIN 1996 2031 EGF-LIKE 8, CALCIUM-BINDING.  
 FT DOMAIN 2037 2070 LAMININ EGF-LIKE.  
 FT DOMAIN 2070 2475 GPS.  
 FT DOMAIN 2475 2678 POLY-LRU.  
 FT DOMAIN 2678 3034 BY SIMILARITY.  
 FT DISULFID 1322 1333 BY SIMILARITY.  
 FT DISULFID 1327 1364 BY SIMILARITY.  
 FT DISULFID 1366 1375 BY SIMILARITY.  
 FT DISULFID 1382 1393 BY SIMILARITY.  
 FT DISULFID 1387 1402 BY SIMILARITY.  
 FT DISULFID 1404 1413 BY SIMILARITY.  
 FT DISULFID 1422 1433 BY SIMILARITY.  
 FT DISULFID 1427 1443 BY SIMILARITY.  
 FT DISULFID 1445 1455 BY SIMILARITY.  
 FT DISULFID 1668 1679 BY SIMILARITY.  
 FT DISULFID 1673 1688 BY SIMILARITY.  
 FT DISULFID 1690 1699 BY SIMILARITY.  
 FT DISULFID 1891 1902 BY SIMILARITY.  
 FT DISULFID 1896 1911 BY SIMILARITY.  
 FT DISULFID 1913 1922 BY SIMILARITY.  
 FT DISULFID 1926 1937 BY SIMILARITY.  
 FT DISULFID 1931 1949 BY SIMILARITY.  
 FT DISULFID 1951 1960 BY SIMILARITY.  
 FT DISULFID 1968 1981 BY SIMILARITY.  
 FT DISULFID 1983 1993 BY SIMILARITY.  
 FT DISULFID 2000 2015 BY SIMILARITY.  
 FT DISULFID 2002 2018 BY SIMILARITY.  
 FT DISULFID 2020 2030 BY SIMILARITY.  
 FT MOD\_RES 1681 1681 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 1904 1904 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 793 793 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1129 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1228 1228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1264 1264 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1302 1302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1591 1591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1655 1655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1994 1994 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2118 2118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2137 2137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2144 2144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2155 2155 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2160 2160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2452 2452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2538 2538 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3034 AA; 330477 MW; EFF38180AF5ED8A8 CRC64;  
 Query Match 57.7%; Score 8974.5; DB 1; Length 3034;  
 Best Local Similarity 57.3%; Fred. No. 0;  
 Matches 1730; Conservative 429; Mismatches 648; Indels 210; Gaps 39;  
 1 MRSPTGVPLPFPPLPLLLLLLPPLIGQVGCRCISGRGSS-----S 48  
 124 LRSNAGALRSP-----AVRSVPLGDL--CFPAAGGAASLTSVLEAITNPA 172  
 49 GACAPN-----GWC--PSSASINMLYTSRCRDAGTELGLVPHDGLRWCPSESA 99  
 173 CSCPPVAGTGCRRGPCLPFGSABELRLCALGRAGA-----VWV----- 213  
 100 HPLPAPGCPMSCLLDIGHLSPQGLTLPEHPCAKAPRLRCQCKLAQAPLRA 159  
 214 -----ELVIQATSGTPSESPV-SPL---LNLSP--RAG 243  
 160 --ERSPESLGRRKKNVTAPQFPSPYQATVPENQAPTPVASIRATDPDEGAGRL 217  
 244 VVRS-----RGGSSSTSPQPLPSYQVSPENEPAGTAVIELRAHDPDEGAGRLS 296  
 218 YTMALPDRSNQFSLDPVTGAVTAEELDRKTSYVFRVYADHGMPPRSALATLT 277  
 297 YQWEALPDRSNQFSLDPVTGAVTAEELDRKTSYVFRVYADHGMPPRSALATLT 356  
 278 LVTDDNDHPVPRQOQYKESLRNLEVGREVTVAATGDAPNANILRLRGSGSPS 337  
 357 TVSDNDHSPVPRQOQYKESLRNLEVGREVTVAATGDAPNANILRLRGSGAGS--- 413  
 338 EYFELIDPRSGVLRTRGPVDRREVEESYQLTVEASDQGRDPGRSTTNAVELSVEDNDNAP 397  
 414 -VEIDARSGVTRAVVDREBAEYQLLVEANDQGRNGLSASATVHIVVEDENDNP 472  
 398 QSEKRYVQVQEDVTGAPVLRVITASDQKSNVHVSINSGNARQGYLDAQGLAD 457  
 473 QSEKRYVQVQEDVAVNTAVTRVQATDRDQGNIAHVSIVSGNKGFPYHLSJSGSLD 532  
 458 VVSPFLDYETTKETLRVRAQDGGRPPLSNVSGLVTVQVLDINDNAP1FSTPQATVLS 517  
 533 VINPLDFEALIRYTLIRAKQDGGRPPLINSGLVSVQVLDVNDNAP1FSSPPQAVL 592  
 518 VPLGYLVLFVQALIDADAGNARLELYLAGVH-----DPEFTINNGTG 560  
 593 VPLGSHVLIQAVNDADAGENARLYQLVDTASTIVGSSVDSNPASADPFOJHNSG 652  
 561 KTSVAEELDRREVDYPSRGVEARDHGTPLATZASVSUTVLVNDNPPFTQPEYVRN 620  
 653 WITVCAELDRREVEHYSFGVEZAVDHGSPMSSASVSITVLVNDNPPFTQPEYELRN 712  
 621 EDAAVGTSVVTVSAVDRDAHVSYYTQITSGNTRNFPSTISQSGGGLVSLALPLDYLERQ 680  
 713 EDAAVSSVTLTRARDRDANSVITQITGNTNRNFPALSSQGGGLITLALPLDYQERQ 772  
 681 YVLAVTASDGTRODTRQIVVNTVDANTHRPFQSSHVTVNVEDRPAGTTVVLISATDE 740  
 773 YVLAVTASDGTSHRQVIVNTDANTHRPVQSSHYTVSEDEPVGSTISATISDE 832  
 741 TGENAITYPMEDSITPOFRLDADTGAVTQALDYEDQVSYLATARPNGIPQSDTTY 800  
 833 TGENAITYVLDPVPOFRLDADTGAVTQALDYEDQVSYLATARPNGIPQSDTTY 892  
 801 LELIVNDVNDNAPQFLRDSYQSVVEDVPFTSVLQISATDRDGLNGRVFTFGGDDG 860  
 893 LELIILDANDNAPRFLRDPYQSVVEDAPPSVSVQVASTDRDGLNGRLVLTFGGDDG 952  
 861 DGDFTYESTSGVTRTLRLDRENVAYVLRAYAVDKMP-PARTPMEVTVTVLVNDNPP 919  
 953 DGFYFTEPTSGVIRTORRLDRENVAYVLMALAVRGSPNPLSASGIVSVTVLDINDNP 1012  
 920 VFEQDFDVFEVENSIGLAVARVATDPDEGNAQIMQIYEGNAPYFOLDIFSGEJT 979  
 1013 VFEKDELLEFVENSFVGSVVARIRANDPDBGNMOIITQIVEGNPFVFDLISGLLR 1072

QY 980 ALVDLDEDEPEYLVITQATSAPLVSAATVHVLLENDNDPVLGNFEILFNNVYNNRSS 1039  
 DB 1073 ALVELPEFVARDVYLVOATSAPLVSAATVHVLLENDNDPVLGNFEILFNNVYNNRSS 1132  
 QY 1040 SPFGAIGRAPADPDSDLSITYSFERGNELSLVLNASTGELKLSALNNRPLEKINS 1099  
 DB 1133 SPFGAIGRAPADPDSDLSITYSFERGNELSLVLNASTGELKLSALNNRPLEKINS 1192  
 QY 1100 VLVSVDGSHVTAOCALRVITITDEMLTHSITLRLDEMSPERPLSLGLFIOAAVATAT 1159  
 DB 1193 VVSVDGSHVTAOCALRVITITDEMLTHSITLRLDEMSPERPLSLGLFIOAAVATAT 1252  
 QY 1160 PPHVAVVFNQRTDAPAGHILNVSLSVGPDPGGPPLPSBDIOERLYNRSLLTAT 1219  
 DB 1253 TKXDIFFVFNQNTDV--SSNINLVTFESALLPGGTG--RFFPSBDIOERLYNRSLLTAT 1309  
 QY 1220 SAORVLPFDNICTLRPCENYMRVSVLRFDSAPFLASSVLPRPHVGGIRCCPCPG 1279  
 DB 1310 SAORVLPFDNICTLRPCENYMRVSVLRFDSAPFLASSVLPRPHVGGIRCCPCPG 1369  
 QY 1280 FTGDYCEYEDVLCYSRCPGPHGRCSRREGYTCLRDYTGHEHCEVSARSGRCTPGVCKN 1339  
 DB 1370 FTGDYCEYEDVLCYSRCPGPHGRCSRREGYTCLRDYTGHEHCEVSARSGRCTPGVCKN 1429  
 QY 1340 GGTGVNLLVGGFPCDCPSGDFEKPVCQVTRSPHASFITRGLRORFHTLALSPATYE 1399  
 DB 1430 GGTGVNLLVGGFPCDCPSGDFEKPVCQVTRSPHASFITRGLRORFHTLALSPATYE 1489  
 QY 1400 RDLGLLYNGRPNKHPVLALEVOEQVOLFESAGESTTVSPPVPGVSGQCHHTYQLK 1459  
 DB 1490 RDLGLLYNGRPNKHPVLALEVOEQVOLFESAGESTTVSPPVPGVSGQCHHTYQLK 1549  
 QY 1460 YNPKLLGQTLPGQPSBOKVAVTVVDCDTGVALARFEGSVLGANSCAAQGTQSGKSLDU 1519  
 DB 1550 YNPKLLGQTLPGQPSBOKVAVTVVDCDTGVALARFEGSVLGANSCAAQGTQSGKSLDU 1609  
 QY 1520 TGPLLLGGVVDLRESPPVNRKROVYGCERNIQUVSRHIDMADFLANNGTGCPAKNVC 1579  
 DB 1610 TGPLLLGGVVDLRESPPVNRKROVYGCERNIQUVSRHIDMADFLANNGTGCPAKNVC 1669  
 QY 1580 SNTCHNGTCVNOMDAFSCPCPLGFGKSCAOMKANDPHLGLSSILVAMHGLSLPIQPMY 1639  
 DB 1670 SNTCHNGTCVNOMDAFSCPCPLGFGKSCAOMKANDPHLGLSSILVAMHGLSLPIQPMY 1729  
 QY 1640 LSLMFTROADGVLLQAITRGRSTITLQLEBGMVLSVBEGTGLQASSLRLEPRANDGM 1699  
 DB 1730 LSLMFTROADGVLLQAITRGRSTITLQLEBGMVLSVBEGTGLQASSLRLEPRANDGM 1789  
 QY 1700 HHAQLALASGGRGH-----AIIISPDYGOGRABGNLGPRLHGLHSNITVVGIRPGAGV 1754  
 DB 1790 HHAQLALASGGRGH-----AIIISPDYGOGRABGNLGPRLHGLHSNITVVGIRPGAGV 1848  
 QY 1755 ARGFRCGLQGVRSVDPBGVNSLDPHGESINVEQCSLPPDPCSNPCANVSICSDNMS 1814  
 DB 1849 ARGFRCGLQGVRSVDPBGVNSLDPHGESINVEQCSLPPDPCSNPCANVSICSDNMS 1908  
 QY 1815 YSCSCDPGYGDNCTVNCIDNLPCEHOSVCTRKPSAHGYTCECPVYLAGYCESTRIDPC 1874  
 DB 1909 YSCSCDPGYGDNCTVNCIDNLPCEHOSVCTRKPSAHGYTCECPVYLAGYCESTRIDPC 1968  
 QY 1875 PRGMWGHFTGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSPYCLLDCYCYTGSLSR 1934  
 DB 1969 PRGMWGHFTGPCNCDVSKGFPDCKNTSGECHKENHTRPPGSPYCLLDCYCYTGSLSR 2028  
 QY 1935 VCPBEOGQCECKPGVIGROCDRCNDPFAVTTNGCENVDSCRAIEAGIMPRTRFGLP 1994  
 DB 2029 VCPBEOGQCECKPGVIGROCDRCNDPFAVTTNGCENVDSCRAIEAGIMPRTRFGLP 2088  
 QY 1995 AAAPCEKSGFATVRRCHDERGMLPMLFNCTITTSBILKGFAERLQRNESGLDSGRSOQ 2054  
 DB 2089 AAAPCEKSGFATVRRCHDERGMLPMLFNCTITTSBILKGFAERLQRNESGLDSGRSOQ 2148

QY 2055 LALLANATQHTAGYFGSDVKAVALQATRLAHSTORGFGLSANQDVHFTENTLAVGSA 2114  
 DB 2149 LALLANATQHTAGYFGSDVKAVALQATRLAHSTORGFGLSANQDVHFTENTLAVGSA 2208  
 QY 2115 LLDTPANKRWEILQOTEGTATLLOHYEAYASALAKONMHTLSPPTIYTPNIVISVRL 2174  
 DB 2209 LLDTPANKRWEILQOTEGTATLLOHYEAYASALAKONMHTLSPPTIYTPNIVISVRL 2268  
 QY 2175 DKGNPAGAKLPRYEMALRGQPDLETTVLIPSSVF-----RNPVYR-----PAGE 2221  
 DB 2269 DKGNPAGAKLPRYEMALRGQPDLETTVLIPSSVF-----RNPVYR-----PAGE 2328  
 QY 2222 GEAQEBEELARQORHPELSQGBAASVITTYTLAGLPHNYDPDKRSIYVRRPIINTP 2281  
 DB 2329 GEAQEBEELARQORHPELSQGBAASVITTYTLAGLPHNYDPDKRSIYVRRPIINTP 2388  
 QY 2282 VVISVYHDBELLPPALDPVTVQORLLTERTRTPICYFMNHSILVSGTQMSARGCEV 2341  
 DB 2389 VVISVYHDBELLPPALDPVTVQORLLTERTRTPICYFMNHSILVSGTQMSARGCEV 2448  
 QY 2342 VFRNESHVSCQCNHMTSPAVLMDVSRRENGIILPKITLYVALGYTALALTFPFLTL 2401  
 DB 2449 VFRNESHVSCQCNHMTSPAVLMDVSRRENGIILPKITLYVALGYTALALTFPFLTL 2508  
 QY 2402 RILRSNQGIRNLTALGLAQVFLGIMQADLPACTVIALHLFVLYCTFSMALTEA 2461  
 DB 2509 RILRSNQGIRNLTALGLAQVFLGIMQADLPACTVIALHLFVLYCTFSMALTEA 2568  
 QY 2462 LHLRALTRVRDVTNTPRPFYMLGWPAPFTGLANGIDPEGYGNPDCMTSTYDTLW 2521  
 DB 2569 LHLRALTRVRDVTNTPRPFYMLGWPAPFTGLANGIDPEGYGNPDCMTSTYDTLW 2628  
 QY 2522 SPAGPVAFAVSMSVLYLAAAPASCAORQGEKKGPVSGLOPSFAVLLLSATVLLAL 2581  
 DB 2629 SPAGPVAFAVSMSVLYLAAAPASCAORQGEKKGPVSGLOPSFAVLLLSATVLLAL 2688  
 QY 2582 SVNSDTLLFHYLPATCNCIQGFIFLSYVLSKEYRAK-LACSRKPSPDALTTKST- 2639  
 DB 2689 SVNSDTLLFHYLPATCNCIQGFIFLSYVLSKEYRAK-LACSRKPSPDALTTKST- 2748  
 QY 2640 LHSYNCSBPVADG--RLYQVPGDSAGLSHSRSBGKSPSVIIFPLRESMLNQCQGP 2697  
 DB 2749 LHSYNCSBPVADG--RLYQVPGDSAGLSHSRSBGKSPSVIIFPLRESMLNQCQGP 2798  
 QY 2698 GLADPGSLPL-EGQPOQHDPTDSDSLILBDDQGSYASTHSDBSEEBEESBAAP 2756  
 DB 2799 GLADPGSLPL-EGQPOQHDPTDSDSLILBDDQGSYASTHSDBSEEBEESBAAP 2853  
 QY 2757 GEQWDSLLGPGBERLPLHSTPK-DGPGPGKAPMPGD--FGTTAKE----- 2800  
 DB 2854 GEQWDSLLGPGBERLPLHSTPK-DGPGPGKAPMPGD--FGTTAKE----- 2903  
 QY 2801 -----SSGNAPBERLRENDALSRBGSGLPLGSSAQP---HGKILKKCL--PTIS 2848  
 DB 2904 -----SSGNAPBERLRENDALSRBGSGLPLGSSAQP---HGKILKKCL--PTIS 2958  
 QY 2849 EK--SLLRLPLQCTGSSRGCS-SASESGRGGPP-----RPPPOSIOQLGNGVPI 2898  
 DB 2959 EK--SLLRLPLQCTGSSRGCS-SASESGRGGPP-----RPPPOSIOQLGNGVPI 3015  
 QY 2899 AMSITAGVDEDSGSE 2915  
 DB 3016 AMSITAGVDEDSGSE 3032

RESULT 5  
 CLRL HUMAN  
 ID CLRL HUMAN STANDARD; PRT, 3014 AA.  
 AC Q9NY06; Q9Y722; Q9BM05; Q9Y506; Q9Y526;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Caderlin Bgf IAG seven-pass G-type receptor 1 precursor (Flamingo

DE homolog 2) (hpm12).  
GN CELSRI OR CDH9 OR FMT2.  
OS Homo sapiens (human).  
OC Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eumetazoa; Buthozoa; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20205259; PubMed=10716726;  
RA Wu O., Maniatis T.;  
RT "Large exons encoding multiple ectodomains are a characteristic  
RT feature of protocadherin genes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
RA Clamp M., Smith L.J., Alincough R., Almeida J.P., Babbage A.K.,  
RA Baggsley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,  
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
RA Conroy D., Corby N.R., Corvill G.J., Cox A.V., Davis J., Dawson B.,  
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
RA Hall S.E., Jones M.C., Kershaw J., Kimberley A.M., Lloyd D.M.,  
RA Hunt S.E., James G.C., Kershaw J., Kimberley A.M., Lloyd D.M.,  
RA Laird G.K., Longford C.P., Leverha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,  
RA McElroy J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.C.T.,  
RA O'Neill C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Rose M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers S., Shmizu N.,  
RA Mnooshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Zhang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chisoso S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Kohling T.,  
RA Schell P., Walker C., Wamley A., Wohlmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L., Wardis E., Waterston R., Wilson R.,  
RA Emanuel H.S., Shaikh T., Kurahashi H., Saito S., Badart M.L.,  
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,  
RA Kim U.J., Shizuya H., Simon M.I., Dumanak J.P., Peyraud M., Kedra D.,  
RA Seroussi E., Franssen I., Tapia I., Brider C.E., O'Brien K.P.,  
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
RA Tiliun Y., Wright H.;  
RT "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495(1999).  
RN  
RP  
RP SEQUENCE OF 624-3014 FROM N.A. (ISOFORM 2).  
RX  
RC TISSUE=Kidney;  
RC MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Haefl N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haefl F.,  
RA Datchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schreier T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Camarini P., Prange C.,  
RA Raha S.S., Loguaillo N.A., Peters G.J., Abranson R.D., Mulhally S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,

RA Villalon D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield Y.S.N., Krzywinski M.T., Skalska U., Skalska D.E.,  
RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC  
CC -1- FUNCTION: Receptor that may have an important role in cell/cell  
CC signaling during nervous system formation.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9NY06-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9NY06-2; Sequence=VSP\_002011, VSP\_002012;  
CC Note=No experimental confirmation available;  
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC -1- SIMILARITY: Contains 9 cadherin domains.  
CC -1- SIMILARITY: Contains 8 EGF-like domains.  
CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.  
CC -1- SIMILARITY: Contains 1 GPS domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF231024; AAF61930.1; -;  
DR EMBL: AL021392; CAB50707.1; -;  
DR EMBL: AL031597; CAB45020.1; -;  
DR EMBL: AL031588; CAB38413.1; -;  
DR EMBL: BC000059; AAH00059.1; -;  
DR Genew; HGNC:1850; CELSRI.  
DR MIM: 604523; -;  
DR HSSP: P00749; IURK.  
DR GO: GO:0016021; C: integral to membrane; NAS.  
DR GO: GO:0008014; F: calcium-dependent cell adhesion molecule ac. ; NAS.  
DR GO: GO:0004930; F: G-protein coupled receptor activity; NAS.  
DR GO: GO:0007399; P: neurogenesis; NAS.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR002126; Cadherin.  
DR InterPro: IPR006209; EGF-like.  
DR InterPro: IPR000832; GPCR\_secretin.  
DR InterPro: IPR001879; homom. receptor.  
DR InterPro: IPR002049; laminin\_EGF.  
DR InterPro: IPR001791; laminin\_G.  
DR InterPro: IPR000203; PKD\_cys\_rich.  
DR Pfam: PF00002; 7tm\_2; 1.  
DR Pfam: PF00028; cadherin; 8.  
DR Pfam: PF00008; EGF; 6.  
DR Pfam: PF01825; GPS; 1.  
DR Pfam: PF02793; HRM; 1.  
DR Pfam: PF00053; laminin\_EGF; 1.  
DR Pfam: PF00054; laminin\_G; 1.  
DR PRINTS: PR00205; CADHERIN.  
DR PRINTS: PR00011; EGF\_LAMININ.  
DR PRINTS: PR00249; GPCRSECRETIN.  
DR SMART: SM00112; CA; 8.  
DR SMART: SM00180; EGF\_Lam; 1.  
DR SMART: SM00303; GPS; 1.  
DR SMART: SM00008; Hornik; 1.  
DR SMART: SM00282; Lamg; 2.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 2.  
DR PROSITE: PS00232; CADHERIN\_1; 7.  
DR PROSITE: PS0268; CADHERIN\_2; 9.

DR PROSITE; PS00022; EGF\_1; 6.  
DR PROSITE; PS0116; EGF\_2; 2.  
DR PROSITE; PS0221; GPs; 1.  
DR PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; FALSE\_NEG.  
DR PROSITE; PS0227; G\_PROTEIN\_RECP\_F2\_3; 1.  
DR PROSITE; PS0261; G\_PROTEIN\_RECP\_F2\_4; 1.  
DR PROSITE; PS01248; LAMININ\_TYRE\_EGF\_1.  
DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
KW Developmental protein; Hydroxylation; Signal; Alternative splicing.  
FT SIGNAL 1 20  
FT CHAIN 21 3014  
FT DOMAIN 22 2469  
FT TRANSMEM 2470 2490  
FT DOMAIN 2491 2501  
FT TRANSMEM 2502 2522  
FT TRANSMEM 2523 2527  
FT TRANSMEM 2528 2548  
FT DOMAIN 2549 2572  
FT TRANSMEM 2573 2593  
FT DOMAIN 2594 2611  
FT TRANSMEM 2612 2632  
FT TRANSMEM 2633 2655  
FT TRANSMEM 2656 2676  
FT TRANSMEM 2677 2683  
FT TRANSMEM 2684 2704  
FT DOMAIN 2705 3014  
FT DOMAIN 246 353  
FT DOMAIN 354 459  
FT DOMAIN 460 565  
FT DOMAIN 566 687  
FT DOMAIN 688 789  
FT DOMAIN 790 892  
FT DOMAIN 893 999  
FT DOMAIN 1000 1101  
FT DOMAIN 1106 1224  
FT DOMAIN 1303 1361  
FT DOMAIN 1363 1399  
FT DOMAIN 1403 1441  
FT DOMAIN 1442 1646  
FT DOMAIN 1649 1685  
FT DOMAIN 1689 1870  
FT DOMAIN 1872 1907  
FT DOMAIN 1908 1946  
FT DOMAIN 1947 1979  
Query Match 56.3%; Score 8754; DB 1; Length 3014;  
Best Local Similarity 55.8%; Pred. No. 0;  
Matches 1703; Conservative 420; Mismatches 687; Indels 240; Gaps 39;  
12 TRPPPLLLLLLLPPPLLDGVPGRSGSRGSSGACAPMGLCPSSASNTMLYTSR 71  
57 TRAPPELL-----DVGGRDRLAGR-RRVSGAGRP-----PLQVRLV----- 93  
72 CRDAGTGLTGLVPHHDLGRVWCPESEAHPLPAPBGGCPMGCRLIGIGHLSPQSKLT 131  
94 ASAPALASRL-----RAKTHP-----GCGARALCCTARLC--GALCF 133  
QY 132 P-----EHPCLAK-----RLRQCQ-----CKLAQAPG- 155  
DB 134 PVPGGCAAGHSLAAPTLLPACRCPRRPRPCGPRICLPGGSVRLTLCALARRAGA 193  
QY 156 LRAG-----ERSPESLGRRRKRNVTAP-----QCPSPYQAT 189  
DB 194 VAVGALAEATAGTSPSPSPPLPPLPAPAGAPARRARAGTSGSLKPPMNYOVA 253  
QY 190 VVENPAGTVPVSLRAIDPDEGAGRLTETMDLFPSSRNOFPSPDPTVGTATAEELDR 249  
DB 254 LEENEAGTLLIQHNAHTYIBGBEBEAVSYVMGLPDBERSRGYFRIDSAIGASTDSVLD 313

QY 250 ETKSTHVFRTAQDHMPRRSALATLLITLITDNDHDPVEEQOBYKESIRENLEVEYTL 309  
DB 314 ETKRHVLAKVKAVDVSTPRSATYITLVVKDNTDHSPIFESSEYRERRENLVEGYTL 373  
QY 310 TYRADGDAPPAANTLLYRLBSGSGSPSEVEFIDPRSGYIRRGPDVREESYQLTVEA 369  
DB 374 TRASDRSPINANRYRLTGGAW-----DVFOLENESSGVSTRVADREEALEYQLVBA 429  
QY 370 SQGGDPGRSTTAAPVLSVEDDNDNAPQFSKRRVYVQRELVGTGAPVLRATSRDRDG 429  
DB 430 NQGRNPGSLATATVYLEVEDENNYPOFSQNTVVQVPEVGTANTAVLRQATDRDG 489  
QY 430 SNAVYHYSIMSGNAGQFYLDAGALDVSPDLDEYETKTYTLRYAODGGRPLSNVSG 489  
DB 490 QNAALHYSILSGNVAGQFYHLSGLIDVYNLDPEDVQKXSLSKAQDGGPPLINSNG 549  
QY 490 LTVVQVLDINDNAPLTVSTPFOATYLSVPGLGYLVLAQALDADGNARLEYRLA---- 545  
DB 550 VVSQVLDVNDNEPLFVSSPFOATVLENPPLGYPVVHTQAVDADGEMARLHYRLVDYAS 609  
QY 546 ----GVGH-----DPEFTINNGTGMISVAELEDREYDFSGYEARDHGTALTA 592  
DB 610 TPLGGSGAGPKKAPATPPPOIHSNGITVCAELDBEHEHYSFGVAVDHSGPMSS 669  
QY 593 SASVSVTVLDVNDNPTFOTPEYTVRLNEDAAVGTSVTVSAVDHDAHSVITYQTSQNT 652  
DB 670 STVSITVLDVNDNPTFOTPEYTVRLNEDAAVGTSVTVLQARBDASVITYQLTGNT 729  
QY 653 RRRPSTQSGGGLVSLALPDYKLEROYVLAATNSDGTQDTAQIVVNTDANTHPVF 712  
DB 730 RRRFALSQRGGGLTTLALPDYKQOQVLAATNSDGRSHAVLINVDTANTHRPVF 789  
QY 733 QSHYTVVWNEDEPAGTTVVVLISATDEPTGENARTTYPMEDSIPQFRIDADGAVTQAE 772  
DB 790 QSHYTVVWNEDEPAGTTVVVLISATDEPTGENARTTYPMEDSIPQFRIDADGAVTQAE 849  
QY 773 LDYEDQSVYTLATARDNGIPQKSDPTYLAILVNDVNDNAPQFLDSYQSGVYEDVPEPT 832  
DB 850 LDYENQVAVYTLTMAQDNGIPQKSDPTYLAILVNDVNDNAPQFLDMPFQGSIFEDAPST 909  
QY 833 SVLQISADDRSGGLNGRFTYPOGGDDDDGPVYESTSGIYRTLRRLRENAQVYLAAY 892  
DB 910 SILQVSADDRSGGPRGLLYTFOGDDDDGPVYESTSGIYRTLRRLRENAQVYLAAY 969  
QY 893 AVDKGMP-PARTPMEVTVTVLDVNDNPTFOTPEYTVRLNEDAAVGTSVTVSAVDHDAHSVITYQTSQNT 951  
DB 970 AVDKGMP-PARTPMEVTVTVLDVNDNPTFOTPEYTVRLNEDAAVGTSVTVSAVDHDAHSVITYQTSQNT 1029  
QY 952 TNAQIMQIVBGNTPVEVQDLPFSGELTALVDLDYEDRPEYVLYQATSAPLVSRATVAV 1011  
DB 1030 PNAQIMQIVBGNTPVEVQDLPFSGELTALVDLDYEDRPEYVLYQATSAPLVSRATVAV 1089  
QY 1012 RLIDRNDNPTPLGAPFELLFNNYVNNRSSPFGCAIGRPADPDIDSLTYSFERNELS 1071  
DB 1090 LLDVQNDNPTPLGAPFELLFNNYVNNRSSPFGCAIGRPADPDIDSLTYSFERNELS 1149  
QY 1072 LVLNASTGELKSRALDNNRPLEAIVSVSDGASTTAQALRVITITBMLTHSITL 1131  
DB 1150 LLLDPAIGELOLRDNDNPTPLAIMEVSDGHSHTAICTARVITITDMLTNSITV 1209  
QY 1132 RLEDMSPERLSPILGLETOAVATLATPPDVVVFVNYQRTDAPGHIINVSISVGP 1191  
DB 1210 RLENNQKFKLSPILGLETOAVATLATPPDVVVFVNYQRTDAPGHIINVSISVGP 1268  
QY 1289 GVRG--QFSESDIOBOYTLNRKLLTITISQRYVLPFDNICLRBPCENYMCVSLRDS 1326  
DB 1252 SAPFIASSSVLFRPIHPVGLRCRCPGFTGDYCEYEDLTCYSPCGPHGCRREGGYT 1311  
QY 1327 SAPFIASSSVLFRPIHPVGLRCRCPGFTGDYCEYEDLTCYSPCGPHGCRREGGYT 1386  
DB 1312 CLCRDGYTGEHCEVSAASGRCTPGVCNNGTGVNLLVGGFKDCPSGDPEKPYCQVYTRS 1371



```

Db 1387 CEEFEDPTEHCHCVDARSRCANGVCKNGTGVNLLIGFHCPCPGREYERCEYETTS 1446
Qy 1372 FPAHSFETRGRLRQRFHTLALSFATERDGLLNGRFNEKADPALAEVIOQVOLTSS 1431
Db 1447 PFGQSVTRGRLRQRFHTLALSFATERDGLLNGRFNEKADPALAEVIOQVOLTSS 1506
Qy 1432 AGESTTVBPVGVSDQGMATVOLKYYNKPPLLGQTGLPGQSEKQAVVTVGCDTGV 1491
Db 1507 AGESTTVBPVGVSDQGMATVOLKYYNKPPLLGQTGLPGQSEKQAVVTVGCDTGV 1566
Qy 1492 ALRFGSVLNGVCAAGTGGSKSLDLNCPGLLLGVPDLPSFPFRMOPGCMRNLQY 1551
Db 1567 AARFGDIGNVSCAAGTGGSKSLDLNCPGLLLGVPDLPSFPFRMOPGCMRNLQY 1626
Qy 1552 DSRHIMADFIANGTVPGCPACKNVCDSNTANGTCVNMOMDFAFSGCEPLRGSGCAQ 1611
Db 1627 DGNVMDAGFIANGTVPGCPACKNVCDSNTANGTCVNMOMDFAFSGCEPLRGSGCAQ 1686
Qy 1612 EMANPQHLGSSLVANHGSLPISQPMYLSLMFRTQADGVLLQAITRGRSTTTQLRBA 1671
Db 1687 AMPHPOLFSGESVVSNDLNIISVPMYLGIMFTRKEDSVLMEATSGGPTSPRLQILNN 1746
Qy 1672 HWMLSEGVGLQASSLRBERGANDDMMHQAOL---GASGPRGHAL-LSPPYGGQRAE 1727
Db 1747 YLQPFVSHGSPDVSVMLSGLRTDGRMHLLLELNKXEDSEMKHLVTMTLDYGMQDK 1806
Qy 1728 GNLGPRHLGHLNITVGGIPGPAGVAFRGCLQGVRSVDPREGNSLDPBSGSIYV 1787
Db 1807 ADIGMPLGLTVASVVGASBKVVRGFRGCMGVNMGITPTVATLNNMNAKAVY 1866
Qy 1788 BQGSCLPDPGDSNCPANSTYCSNDMDSYSCSDPGYEDNCTNVCDLNPCEHOSVCTR 1847
Db 1867 KQGCDDVDDCTSSPCPNRSKCHAMEDYSCVCKGYLGINCVDACHLNCENMGACVRS 1926
Qy 1848 SAPHYTCEPRMYLGPYCESTRIDOPCRGMGHPTCGPCNCDVSGSPBDCKTSGE 1907
Db 1927 GSPQGVCEGPRHGYCENKLDLPCPRGMGNPVCSPGHCAVSGPBDCKTNGQCO 1986
Qy 1908 CKENHYRPGSPYCLLDCYPTGSLRVCDPBDQCPCKKVGIGROCDRCNPFPAVTN 1967
Db 1987 CKENHYKLAQDTCPCDPCPHGSHRTCDMAGQCAKRGVIGROCDRCNPFPAVTN 2046
Qy 1968 GCEVNTDSCPRALFAGIWMPTRTFGRLPAAAPCPKSGFRTAVNCHDEHGMPLNFC 2027
Db 2047 GCEVNTDSCPRALFAGIWMPTRTFGRLPAAAPCPKSGFRTAVNCHDEHGMPLNFC 2106
Qy 2028 ITSEELKGAERLQHNESGLDSRQQLALLLNATQHTAGVGSVDKVAYOLATLLAH 2087
Db 2107 ISFVDLRANMEKLSRNBSTOVGABALQVRLASATQHTGLTGNVTRTAYQLGHV 2166
Qy 2088 ESTORFGSLSATODVHFTENLTVGSALLDTANKRMELIQTEGGTAMLLQHYEAYASA 2147
Db 2167 ESNQOGFDLAATODADHEDVHISGALLAPATRAMBOQISEGGTAQLRLRLEBYFN 2226
Qy 2148 LAQNMHTTISPTTYTPTNIVISVRLDKNPNAGALPRYALRGSPDPLETTVLPBS 2207
Db 2227 VARNVRYTRPVIYATNMILAVDFDKNFPGARPRDTHIERPRELBSVSFPAD 2286
Qy 2208 VFR-----ETPPVVRPAG-----PGEAQBEBELARQRHPELSQGAVALSVIY 2254
Db 2287 FFRPPEKSGPLRPAQRRTTPOTTRPGPETERBAPISRRRRPDDGCPAVALVIY 2346
Qy 2255 LAGLLPHNTYDPRKSLRVPKRPINTPVVSVIHDEBELPRALDKPVTVQFLLTER 2314
Db 2347 LQGLLPRYDPRRSLRLPRPIINTPVVSTLVSEBAPLPRLEBPVLEFALLEVER 2406
Qy 2315 TKRICYPMNHSILVSGSGSAGCEVFRNESHVSQCNHMTSPAVLMDVSRNGE 2374
Db 2407 TKRVCYPMNHSILVSGSGSAGCEVFRNESHVSQCNHMTSPAVLMDVSRNGE 2466
Qy 2375 PLKTLVVALGVTLAALLTFPFLTLRLIRSNQHGIRMLTAAIGLAQVFLGIGNAD 2434

```

```

Db 2467 PIKIVTAVALSLSALLVAVFLVSLVMLRSNHSIKHLAVALFLSOLVFIIGINTE 2526
Qy 2435 LPPACTVAILLHFLYLCFSPMALLLEALHLYALTEVDVNTGPMRPFYMLGMGPAT 2494
Db 2527 NPFCLTVAILLHFLYLCFSPMALLLEALHLYALTEVDVNTGPMRPFYMLGMGPAT 2586
Qy 2495 GLAVGLDEPGYGNPFCWLSIYDVLINSFAPVAVASVMSVFLYTLAARASCAORGE 2554
Db 2587 GLAVGLDEPGYGNPFCWLSIYDVLINSFAPVAVASVMSVFLYTLAARASCAORGE 2646
Qy 2555 KKGPSVGLPSPAVYLLLSATMTALLSVNSDILLFHTLPATCNIOGPFIPLSVYLSK 2614
Db 2647 KKGPSVGLPSPAVYLLLSATMTALLSVNSDILLFHTLPATCNIOGPFIPLSVYLSK 2706
Qy 2615 EVRKALK-LACSRKSPDPALTTKST-LTSSYNCSPADG--RLYQPYGSAGSLHSTS 2670
Db 2707 EVRKALKVGLGRKHLHDSATTRATLLTRSLNCMTTGDGPDMLRTLGSTASLDIV 2766
Qy 2671 RS-----GKSQPSYIPPLLRBSALNPGQPPGLDPSLFLGQDQO 2713
Db 2767 RDEGLOKLGVSSGLVRGSHGPRDASIMPRSCD-----PFG----- 2802
Qy 2714 HDPDTPSDLSLBDQSGSYASTHSSDSEEBEERBAAPPGSGWDSLLGPABRLP 2773
Db 2803 --HDSDSSELSL--DEQSSSTASHSSDSEDDGVABEK-----WDPARG-----A 2845
Qy 2774 LHSTPK-DGSGPCPKAPMPDFTGTAKESSGNGAP-----EERLRNGDAL 2818
Db 2846 VASTPKGAVANHVAPMGPDGSLAESDSEDPGKRLAVETKVSYLEHREGSGHRY 2905
Qy 2819 SREGLPLPGSSAOP--HGILKKCL---PTISKS--SLRLPLEOCTGSSRGS 2869
Db 2906 PDBSGGAARLASQPPQKRGILNKATYPPPLTLTQTLKGRREKLACGSPSSR 2965
Qy 2870 ASESGRGP-----PPRPPRQSLQOLNCPMIMSIRAGTVDBSSGS 2915
Db 2966 TSSLSGGPDCAITYKSPRERGRDLNGV--ANVATGSAQDGSSE 3012

RESULT 6
CLR3 HUMAN
ID CLR3 HUMAN STANDARD; PRT; 3312 AA.
AC 09NTQ7; 075092;
DT 28-FEB-2003 (rel. 41, Created)
DT 28-FEB-2003 (rel. 41, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Cadherin EGF lag seven-pass G-type receptor 3 precursor (Flamingo homolog 1) (hFml1) (Multiple epidermal growth factor-like domains 2) (Epidermal growth factor-like 1).
GN CELSR3 OR CDHF11 OR FMI1 OR EGFL1 OR MEGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

```

CC -1- SIMILARITY: Contains 9 cadherin domains.  
 CC -1- SIMILARITY: Contains 8 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF231023; AAF61929.1; --  
 CC EMBL; AB011536; BAA32464.1; --  
 CC HSSP; P00740; 1BDM.  
 CC Genew; HGNC:3230; CELSR3.  
 CC MIM; 604264; --  
 CC GO; GO:0005198; F:structural molecule activity; NAS.  
 CC InterPro; IPR000152; Asx\_hydroxyl.  
 CC InterPro; IPR002126; Cadherin.  
 CC InterPro; IPR000742; EGF\_2.  
 CC InterPro; IPR001881; EGF\_Ca.  
 CC InterPro; IPR006209; EGF\_like.  
 CC InterPro; IPR000832; GPCR\_receptor.  
 CC InterPro; IPR001879; hormn\_receptor.  
 CC InterPro; IPR006210; IEGF.  
 CC InterPro; IPR001791; laminin\_G.  
 CC InterPro; IPR000203; PKD\_cys\_rich.  
 CC Pfam; PF00002; 7tm\_2; 1.  
 CC Pfam; PF00028; cadherin; 9.  
 CC Pfam; PF00008; EGF; 6.  
 CC Pfam; PF01825; HRM; 1.  
 CC Pfam; PF02793; HRM; 1.  
 CC Pfam; PF00054; laminin\_G; 2.  
 CC PRINTS; PR00205; CADHERIN.  
 CC PRINTS; PR00011; EGF\_LAMININ.  
 CC PRINTS; PR00249; GPCRSECRETIN.  
 CC SMART; SM00112; CA; 9.  
 CC SMART; SM00181; EGF; 6.  
 CC SMART; SM00303; GPS; 1.  
 CC SMART; SM00008; Hormr; 1.  
 CC SMART; SM00282; Lamg; 2.  
 CC PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 CC PROSITE; PS00232; CADHERIN\_1; 7.  
 CC PROSITE; PS00268; CADHERIN\_2; 8.  
 CC PROSITE; PS00022; EGF\_1; 6.  
 CC PROSITE; PS01186; EGF\_2; 4.  
 CC PROSITE; PS00221; GPS; 1.  
 CC PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 CC PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 CC PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 CC PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 CC PROSITE; PS00265; LAMININ\_TYPE\_EGF; 1.  
 CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 CC G-protein coupled receptor; Transmembrane; Glycoprotein;  
 CC EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;  
 CC Developmental protein; Hydroxylation; Signal.  
 CC SIGNAL 1 32  
 CC CHAIN 33 3312  
 CC CDS 33 3312  
 CC DOMAIN 33 2540  
 CC TRANSMEM 2541 2561  
 CC DOMAIN 2562 2572  
 CC TRANSMEM 2573 2593  
 CC DOMAIN 2594 2601  
 CC TRANSMEM 2602 2622  
 CC DOMAIN 2623 2643  
 CC TRANSMEM 2644 2664  
 CC DOMAIN 2665 2681  
 CC TRANSMEM 2682 2702

FT DOMAIN 2703 2725  
 FT TRANSMEM 2726 2746  
 FT DOMAIN 2747 2753  
 FT TRANSMEM 2754 2774  
 FT DOMAIN 2775 3312  
 FT DOMAIN 3312 433  
 FT DOMAIN 434 545  
 FT DOMAIN 546 651  
 FT DOMAIN 652 756  
 FT DOMAIN 757 858  
 FT DOMAIN 859 961  
 FT DOMAIN 962 1067  
 FT DOMAIN 1068 1169  
 FT DOMAIN 1170 1265  
 FT DOMAIN 1266 1375  
 FT DOMAIN 1376 1471  
 FT DOMAIN 1472 1514  
 FT DOMAIN 1515 1719  
 FT DOMAIN 1720 1758  
 FT DOMAIN 1759 1944  
 FT DOMAIN 1945 1982  
 FT DOMAIN 1983 2020  
 FT DOMAIN 2021 2053  
 FT DOMAIN 2054 2090  
 FT DOMAIN 2091 2131  
 FT DOMAIN 2132 2259  
 FT DOMAIN 2260 1379  
 FT DOMAIN 1380 1421  
 FT DOMAIN 1422 1432  
 FT DOMAIN 1433 1450  
 FT DOMAIN 1451 1459  
 FT DOMAIN 1460 1470  
 FT DOMAIN 1471 1500  
 FT DOMAIN 1501 1513  
 FT DOMAIN 1514 1737  
 FT DOMAIN 1738 1746  
 FT DOMAIN 1747 1757  
 FT DOMAIN 1758 1961  
 FT DOMAIN 1962 1970  
 FT DOMAIN 1971 1981  
 FT DOMAIN 1982 1996  
 FT DOMAIN 1997 2008  
 FT DOMAIN 2009 2019  
 FT DOMAIN 2020 2040  
 FT DOMAIN 2041 2052  
 FT DOMAIN 2053 2074  
 FT DOMAIN 2075 2077  
 FT DOMAIN 2078 2089  
 FT DOMAIN 2090 1963  
 FT DOMAIN 1964 632  
 FT DOMAIN 633 847  
 FT DOMAIN 848 1182  
 FT DOMAIN 1183 1222  
 FT DOMAIN 1223 1317  
 FT DOMAIN 1318 1327  
 FT DOMAIN 1328 1649  
 FT DOMAIN 1650 1713  
 FT DOMAIN 1714 1770  
 FT DOMAIN 1771 2053  
 FT DOMAIN 2054 2177  
 FT DOMAIN 2178 2196  
 FT DOMAIN 2197 2386  
 FT DOMAIN 2387 2474  
 FT DOMAIN 2475 2506  
 FT DOMAIN 2507 2158  
 FT DOMAIN 2159 3312

Query Match 50.4%; Score 783.5; DB 1; Length 3312;  
 Best Local Similarity 52.8%; Pred. No. 0;  
 Matches 1558; Conservative 421; Mismatches 801; Indels 173; Gaps 44;

QY 41 GSRGSSGACAPMGWLCPPSSASNLMTYTRCRDAGTELTG--HLVPHDGLRVWCPS 97

```

194 GSRRRVTAACGCELMATGSKGQ-----ERATTSKAKRTAPRNNCLFGASG-----240
QY 98 EAHIFLPAPBGCWPCRLIGIGHLSPOGKLTLPBHPCLKAPRLKOS---CK-LAQA 153
Db 241 -----SGPBLDSAPRTARTAPASGS-APRESRTPEBAP-----KMSRGLFCRFLPQR 230
QY 154 PG-----LRAGSRPEESLGGRRKNVNTAPQPPSYQATVPENQAPTPVASLRAT 206
Db 291 PGPRPPLPARPEARKTYSANRAFRRAANHPQPPQYNYQTLVPEBAAGTAVLRVAQ 350
QY 207 DPBGEAGLEXTMDALPDSRSNQFSLDPVTGAVTAAELDPKSTKTHFRYTAQDHM 266
Db 351 DPDAEGEGRVYSLAALMNSRSLFSLIDPQSGLIRTAALDRSMERHILRYTAQDHG 410
QY 267 PRSALATLTLVYTDNDHPVEEQOEKESIRENLEVEYEVLTARATDGDAPNNAILY 326
Db 411 PRSATMTAVATVADRNDHSPVEQOAYRETLRENVBEGPIIQLRATDDBAPNNILRY 470
QY 327 RLL--EGSGSPSEVFEIDPPSGVIRTRGPVDEEVESYQUTVASDQGRPGPRSTAA 384
Db 471 RFGVPPARAAAAAFAEIDPRSGLISTSGVREHMESEYELVVEASDQGEPRSRATVR 530
QY 385 VFLSVEDNDNMAPPSSEKRYVVOVREDEVTPGAPVLRYTASDRKGSNAVYHSGMSGAR 444
Db 531 VHTVLDENNDNAPQPSSEKRYVVOVREDEVTPGAPVLRYTADRDNDKAGLVHYNISGNSR 590
QY 445 GQFVLQDTALDVSPLDYETTKCYTLARVADQGRPLSNVSGLVTVQVLDINDNAP 504
Db 591 GHFAIDSLTGEIVQVAPLDPREAREVALRIADQAGRPPLSNVGLASIQVVDINHPI 650
QY 505 FVSTPQATVLESVPLGLVLAHQVADADAGNARLEYRLAGVHDPEPTINNCTWISV 564
Db 651 FVSTPQATVLESVPLGLVLAHQVADADAGNARLEYRLAGVHDPEPTINNCTWISV 710
QY 565 AABLDREVDYFSGRGARHGPALRYTASVSSTVLDVNDNPTFPQPEYTRLMDDA 624
Db 711 SGPLDRSVEHTEFGEARHDSGPLSASAVTVTVLDVNDNPEFTKEXYHRLNDDAA 770
QY 625 VGTSVTVASAVDRDASVITYQUTSNTNRFPSTSGGGGLVSLALPLDYKLERQVLA 684
Db 771 VGTSVTVASAVDRDASVITYQUTSNTNRFPSTSGGGGLVSLALPLDYKLERQVLA 830
QY 685 VTSADGRTQATQIVAVNTDANTRPVQOSHTYVANNEDRPACTTVYLISATDEDTGEN 744
Db 831 LITASDRALHCHCYHINITANTHRPVQSAHVSANEDRPACTTVYLISATDEDTGEN 890
QY 745 ARITYFMEISIPORIRADGAVTTOABLDYEDOVSTLAIATARDNGIPOKSPITYLEIL 804
Db 891 ARITYFMEISIPORIRADGAVTTOABLDYEDOVSTLAIATARDNGIPOKSPITYLEIL 950
QY 805 UNVDNNAPOFLRDSYOGSVYEDVPPTSVLQISATRDSGLNGRVFTEQGDGDDGF 864
Db 951 UNVDNNAPOFLRDSYOGSVYEDVPPTSVLQISATRDSGLNGRVFTEQGDGDDGF 1010
QY 865 IVESSTGIVTLRLDREBNAQVYLAIVADKMPARPATMEVTVYTLVNDNPPVFEQD 924
Db 1011 TIEPTSGIVTLRLDREBNAQVYLAIVADKMPARPATMEVTVYTLVNDNPPVFEQD 1070
QY 925 EFDVFEVENSPIGLAVAVRTATDPBEGTNOQIMQVIEGNIPEYFOLDISGELTALVDL 984
Db 1071 EFDVFEVENSPIGLAVAVRTATDPBEGTNOQIMQVIEGNIPEYFOLDISGELTALVDL 1130
QY 985 DYEDRPVYVLTQATSAPLVSRAIVHRLDRNDPVPULANFELFNANNYTNSSSPFG 1044
Db 1131 DYEDRPVYVLTQATSAPLVSRAIVHRLDRNDPVPULANFELFNANNYTNSSSPFG 1190
QY 1045 AIGRVAPRHPDISLSTYSPERKNEGLVVLNASTGELKSRALDNNRPBALMSVUSD 1104
Db 1191 AIGRVAPRHPDISLSTYSPERKNEGLVVLNASTGELKSRALDNNRPBALMSVUSD 1250
QY 1105 GVHSVTACALRYITITDEMTHSITLRLDEMSPERFLSPLGLFIOAVATLATPPDHV 1164

```

```

1251 GLHSVTAQCVRVITITTELLANSITVRLENMQRFLSPLLGRFLEGVAAVLAAPAEV 1310
QY 1165 VFNVOVOTDADPGHILANVLSVGPBGGGP--PFLPSDLOEJLYNRSILTALSAQ 1222
Db 1311 PIFNIQNDIV- GGVYANVSPALAPRGAGAGAPWPSBELQBLVYRBAALARSLL 1369
QY 1223 RVLPEDDNI CLBPCENTNRCSVLRFDSSAPFLASSVLPPIHVPGLRCRCPPGFTG 1282
Db 1370 DVLPEDDNVCLEBPCENTNRCSVLRFDSSAPFLASSVLPPIHVPGLRCRCPPGFTG 1429
QY 1283 DYCETEVLDI CVSRPCGPHRCRSREBGTCLCRDVTGHHGVASASGCTGVCNNGCT 1342
Db 1430 DYCETEVLDI CVSRPCGPHRCRSREBGTCLCRDVTGHHGVASASGCTGVCNNGCT 1489
QY 1343 CVMILVGFCKDCPCSD- FEKPYCOVTTSPFAHSFITERGLRORHFTALSFATKED 1401
Db 1490 CVDAPNGRCQCPAGANBEGRCVTAANSFPSSFWMRGLRQFHLTSLSPATVQGS 1549
QY 1402 GLLITNGRNEKHDVLALEVOYQVLTFSAGESSTTWSPPVPGVSDQMTVOLKYN 1461
Db 1550 GLLITNGRNEKHDVLALEVOYQVLTFSAGESSTTWSPPVPGVSDQMTVOLKYN 1609
QY 1462 KPFLGQGTLPQGSBQKVAVVTVVGDGDPVALRFGSVLVGNSCAAGTQGSKKSLDLTG 1521
Db 1610 KPFLGQGTLPQGSBQKVAVVTVVGDGDPVALRFGSVLVGNSCAAGTQGSKKSLDLTG 1669
QY 1522 PLLGQVPLPESFVRMRQVPGCKRNLQVDSRHLDMADFIANNGTVPCCPAKKVCDN 1581
Db 1670 PLLGQVPLPESFVRMRQVPGCKRNLQVDSRHLDMADFIANNGTVPCCPAKKVCDN 1729
QY 1582 TCHNGSTCVNOMDASCBCEPLFGGKSCAQEBANQHFAGSLVAMH-GLSLPIQPVYL 1640
Db 1730 TCHNGSTCVNOMDASCBCEPLFGGKSCAQEBANQHFAGSLVAMH-GLSLPIQPVYL 1789
QY 1641 SLMPRTROADGVLTALITRGRSTTLQLRBGMVLSV-BGTGLQASSLLEBRANDGM 1699
Db 1790 SLMPRTROADGVLTALITRGRSTTLQLRBGMVLSV-BGTGLQASSLLEBRANDGM 1848
QY 1700 HNAQIAL--GASGGRHAL--SPDYGOORABGNLAPRLGHLNITVGGI-PCBAGV 1754
Db 1849 HNAQIAL--GASGGRHAL--SPDYGOORABGNLAPRLGHLNITVGGI-PCBAGV 1908
QY 1755 ARGFRGCLQGVAVSDTPEGVNS-LPDSHGESINVGQGSLLPPCCSNPPRANYSYCNMD 1813
Db 1909 ARGFRGCLQGVAVSDTPEGVNS-LPDSHGESINVGQGSLLPPCCSNPPRANYSYCNMD 1966
QY 1814 SYSCSDPGYTDNCTNVCDLAPCEHOSVCTKRPAPRAGYTCECPNYLGPYCETRIDP 1873
Db 1967 SYSCSDPGYTDNCTNVCDLAPCEHOSVCTKRPAPRAGYTCECPNYLGPYCETRIDP 2026
QY 1874 CPGWMMGHPTGCPNCDVSKGFPDPCNTKSGECHKENHYRPPGSPPTCLLDCYPTGSL 1933
Db 2027 CPGWMMGHPTGCPNCDVSKGFPDPCNTKSGECHKENHYRPPGSPPTCLLDCYPTGSL 2086
QY 1934 RUCDEPDGQCPCKEYVIGRQCRCNRPARYTTCENVVYDSCRAIEMGIMWPRRETL 1993
Db 2087 RUCDEPDGQCPCKEYVIGRQCRCNRPARYTTCENVVYDSCRAIEMGIMWPRRETL 2146
QY 1994 PAAAPCKSGPGTAVACHDEHGMPLPMLFNCTSIYFSELKAPARLQKRNESGSGSQ 2053
Db 2147 PAAAPCKSGPGTAVACHDEHGMPLPMLFNCTSIYFSELKAPARLQKRNESGSGSQ 2206
QY 2054 QIALILRNATOTHTAGYFGSDVAVAVOLATRLIAHESYORFGSLATODVHTENILRYGS 2113
Db 2207 QIALILRNATOTHTAGYFGSDVAVAVOLATRLIAHESYORFGSLATODVHTENILRYGS 2266
QY 2114 ALLDPTANKHMEIL-QOTEG--GTAMLOHTEAVASALAOHMRHTYLSPTIYTPNYI 2169
Db 2267 ALLDPTANKHMEIL-QOTEG--GTAMLOHTEAVASALAOHMRHTYLSPTIYTPNYI 2326
QY 2170 SVURLDKGNF--AGALPRYEA--LRGEOPDLATTVILPESVFRETPPVAPAGGEAQ 2225
Db 2327 SVURLDKGNF--AGALPRYEA--LRGEOPDLATTVILPESVFRETPPVAPAGGEAQ 2386

```

```

QY 2226 EBEELARRQRHPELSQGEAVAVIIYRTLAGLPHNYDPDKRSILVPRPIINTEPVSI 2285
DB 2387 STTSVPPPPAPPEPEPGISIIILVYRTLAGLPHNQFQABRRGANLPQNPWNSVVSIV 2446
QY 2286 SYHDBELLPRALDKVYVQFRLTEBETKPICVFWMHSILVSGTSGMSARCEVFN 2345
DB 2447 AVFHGNFRIGLIESISLEFRLQTNANKSCICVQWDPGGLABQHWATADCELVN 2506
QY 2346 ESHVSCQCHMTSPFALMDVSRRE--NGEILPLKTLTYALGVTLAAILTFEFLTLRI 2403
DB 2507 GSHARCRCRTGTGVLMDASPRERLEGDLIAVTHVAVSVAVLTAAILSLIS 2566
QY 2404 LRSNQHRRNLTALAGLQVFLGGINQADLPACTVTAIILHFLYCTSPFMALEALH 2463
DB 2567 LKSNVGHANVAALGVAELFLGIGHKTHNQLVCTAVAILLHFEFLTFAMLFVQGH 2626
QY 2464 LYRALTEVDVNTGPRFYTLMGWYPATITGLVGLDEPGNDPFCMLSTYDTLMSF 2523
DB 2627 LYRMOVPEPNVDRGMRFYHALGWMGPAVLGLAVGLDEGYGNDPFCWISYHEPLWSF 2686
QY 2524 AGPVAFAVMSYFLYTLAARASCA--AORQFEKGPVSGLOPSPAVTLILSLATMILALIS 2582
DB 2687 AGPVVLVIVMGTMTFLAARCTSGTQREA--KTSALTILKSSFLILLVLSASMLFGLIA 2744
QY 2583 VNSDTLLFYLTPATCNCIQGPPFLSYVVLKSKVRKALTLAC--SRKSPDPALTTKSTLT 2641
DB 2745 VHSIILAFYHLAAGLCGLQGLAVLFLCYLMDARAAWMPALGKRAPEARPAAGLGP 2804
QY 2642 SSYNCPSPADGRQLVQ--PYGDSAGSLHSTSRGSKOP-----SYT--PFLIREBSA-- 2689
DB 2805 GAVNYTALFEESGGLRITLGAISTVSSVSARSGRITODDQSGRSTLRNVIVRSGSLAD 2864
QY 2690 -----LNPQGGPPGAG-----DPGSLFLEGGDDQDHPDTSDDLSTLEDDSGSVASTH 2738
DB 2865 HTDHSLOAHAGTDDVAMFPHDAGA-----DSDDSDLSLEESLSLIPSS 2912
QY 2739 SSSSEEEBEEBEEBAAPGEGMSDILGPAERLPLHSTPKGQGPQKAPVPG----- 2792
DB 2913 SEDNERTGRFQPLCRAAQ-----SERLLTH--PKVDGNDLLSYPALGECBA 2960
QY 2793 -----DRGT-----TAKESSGNGAPPERLRKENDALSREGSLGPLGSSAQCHKIL 2839
DB 2961 APCALQTWGSERRIGLDTSKDANNQPPPAL--TSGD-----ETSL-----GRQORRKGL 3011
QY 2840 KKKCLPTISEKSLRLP-----LEQCTGSSRSGSA-----SESGRGAPP 2880
DB 3012 KKR-----LQPLVLPQTRGAPELSKRAATLGRRAVPAASYGRITYAGGCTGLS 3060
QY 2881 RPPPROSLQEOIN 2893
DB 3061 QPASRYSRQOLD 3073

```

```

RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34 (1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC cerebellum, olfactory bulb, cerebral cortex, hippocampus and
CC brain stem.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -1- SIMILARITY: Contains 9 cadherin domain.
CC -1- SIMILARITY: Contains 8 EGF-like domain.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011528; BAA32459.1; -.
DR HSSP; P00740; IEDM.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; hormn_receptor.
DR InterPro; IPR006210; TRGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000203; PRD_cys_rich.
DR Pfam; PF00002; 7tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 6.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; Hormr; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00232; CADHERIN_1; 7.
DR PROSITE; PS00268; CADHERIN_2; 8.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS50221; GPS; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
DR PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
DR PROSITE; PS50227; G_PROTEIN_REC_P2_3; 1.
DR PROSITE; PS50261; G_PROTEIN_REC_P2_4; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 31
FT CHAIN 32 3313
FT FT 32 2538
FT FT 2539 2559
FT FT 2560 2570
FT TRANSMEM 2571 2591

```

PT	DOMAIN	2592	2599	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	2600	2620	3 (POTENTIAL).
PT	TRANSMEM	2621	2641	CYTOPLASMIC (POTENTIAL).
PT	TRANSMEM	2642	2662	4 (POTENTIAL).
PT	TRANSMEM	2663	2679	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	2680	2700	5 (POTENTIAL).
PT	TRANSMEM	2701	2724	CYTOPLASMIC (POTENTIAL).
PT	TRANSMEM	2725	2745	6 (POTENTIAL).
PT	TRANSMEM	2746	2752	EXTRACELLULAR (POTENTIAL).
PT	TRANSMEM	2753	2773	7 (POTENTIAL).
PT	TRANSMEM	2774	3313	CYTOPLASMIC (POTENTIAL).
PT	DOMAIN	317	424	CADHERIN 1.
PT	DOMAIN	425	536	CADHERIN 2.
PT	DOMAIN	537	642	CADHERIN 3.
PT	DOMAIN	643	747	CADHERIN 4.
PT	DOMAIN	748	849	CADHERIN 5.
PT	DOMAIN	850	952	CADHERIN 6.
PT	DOMAIN	953	1058	CADHERIN 7.
PT	DOMAIN	1059	1160	CADHERIN 8.
PT	DOMAIN	1059	1257	CADHERIN 9.
PT	DOMAIN	1366	1424	EGF-LIKE 1, CALCIUM-BINDING.
PT	DOMAIN	1426	1462	EGF-LIKE 2, CALCIUM-BINDING.
PT	DOMAIN	1466	1505	EGF-LIKE 3, CALCIUM-BINDING.
PT	DOMAIN	1506	1710	LAMININ G-LIKE 1.
PT	DOMAIN	1713	1749	EGF-LIKE 4, CALCIUM-BINDING.
PT	DOMAIN	1753	1935	LAMININ G-LIKE 2.
PT	DOMAIN	1937	1972	EGF-LIKE 5, CALCIUM-BINDING.
PT	DOMAIN	1973	2011	EGF-LIKE 6, CALCIUM-BINDING.
PT	DOMAIN	2012	2044	EGF-LIKE 7, CALCIUM-BINDING.
PT	DOMAIN	2046	2081	EGF-LIKE 8, CALCIUM-BINDING.
PT	DOMAIN	2087	2120	LAMININ EGF-LIKE.
PT	DOMAIN	2475	2527	GPS.
PT	DISULFID	1370	1381	BY SIMILARITY.
PT	DISULFID	1375	1412	BY SIMILARITY.
PT	DISULFID	1414	1423	BY SIMILARITY.
PT	DISULFID	1430	1441	BY SIMILARITY.
PT	DISULFID	1435	1450	BY SIMILARITY.
PT	DISULFID	1452	1461	BY SIMILARITY.
PT	DISULFID	1470	1481	BY SIMILARITY.
PT	DISULFID	1493	1491	BY SIMILARITY.
PT	DISULFID	1493	1504	BY SIMILARITY.
PT	DISULFID	1717	1728	BY SIMILARITY.
PT	DISULFID	1722	1737	BY SIMILARITY.
PT	DISULFID	1739	1748	BY SIMILARITY.
PT	DISULFID	1941	1952	BY SIMILARITY.
PT	DISULFID	1946	1961	BY SIMILARITY.
PT	DISULFID	1963	1972	BY SIMILARITY.
PT	DISULFID	1976	1987	BY SIMILARITY.
PT	DISULFID	1981	1999	BY SIMILARITY.
PT	DISULFID	2001	2010	BY SIMILARITY.
PT	DISULFID	2018	2031	BY SIMILARITY.
PT	DISULFID	2033	2043	BY SIMILARITY.
PT	DISULFID	2050	2065	BY SIMILARITY.
PT	DISULFID	2052	2068	BY SIMILARITY.
PT	DISULFID	2070	2080	BY SIMILARITY.
PT	MOD RES	1954	1954	HYDROXYLATION (POTENTIAL).
PT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	838	838	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1173	1173	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1308	1308	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1318	1318	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1640	1640	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1704	1704	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	1761	1761	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2173	2173	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2192	2192	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2382	2382	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2472	2472	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	CARBOHYD	2504	2504	N-LINKED (GLCNAC. . .) (POTENTIAL).
PT	SEQUENCE	3313	359348	AA; 359348 MW; B11DA09517288764 CRC64;

Qy	Query Match	50.1%; Score 7780.5; DB 1; Length 3313;
Db	Best Local Similarity	53.2%; Pred. No. 0;
Db	Matches 1540; Conservative 414; Mismatches 761; Indels 182; Gaps 44;	
Qy	118 GIGGHLSPGKLTLP-----EEHPCAKAPRLRQSGCKLAQAPL--RAGRRSEBSLG---168	
Db	237 GIGGSLAPAVTAAPGAPAPBSRTAPB--RKRSGFLRGRGFLFRPPGPRPGPTGABA 295	
Qy	169 -----GRRKRVNTAPQFPQPSQATVPENQAGTPVASLRAIDPDEBAGLELYTM 220	
Db	296 KRLISTNQARSRRAANRRHPQPPQVNYQTLVPENEAAGTAVLRVAQDDPGSAGLTVSL 355	
Qy	221 DALPFRSNQFPSLDPVTGAVTTAAELDREYKSTVFVNTAODHGMPPRSALATTLTLYT 280	
Db	356 AALMMSRSLFELPSIDPQGLIRTAALDRSMEHRYLVTAADHPSPLSATTTWAVVYA 415	
Qy	281 DPNDDHPVEQOEYVESIRENLEVGXYLVTAATGDAIPNNAILYRLLESGGSP---336	
Db	416 DKNDAHPVEQOYRETRRENVESGYPLQLATATGDAIPNNAILRFLV---GSPAAFT 471	
Qy	337 --SEYFEIDPRSGVTRTGPVDRBEVESYQLTVEASDQGRDGPSTTAAPVLSYEDND 394	
Db	472 AAAAAFEIDPRSGGLSTSGRVDRHESYELVLEASDQGEGRPSATRVHAIITVLEND 531	
Qy	395 NAPPSEKRYVQVEDVTPGAPVLRVTASDRKGSNAVHYSINSNARGQFYIDAOTG 454	
Db	532 NAPPSEKRYVQVEDVTPGAPVLRVTATDOKANGLVHNIISGSRGHFAIDSLTG 591	
Qy	455 ALDVSPIDYETTKRYTLRVADQGRPLASVGLVYQVLDINDNAPIPSTPFOATV 514	
Db	592 EIQVNAIPDFEAKERYALIRAQDNGRPLSNITGLASIQQVDINDHSIFVSTFPQSV 651	
Qy	515 LESVPLGYLVAVQAIDADAGDAELRYRLAGVGHDPFTINNGTMSVAAELDREVD 574	
Db	652 LEMALGHSHVHIQAVDADHGENSLRLEYSLGVASDFTFVINSAGWVSQGPLDRESVE 711	
Qy	575 FYSFGVEARDHGTPLTASASVTVLVNDNNPFTTQPEYTVRLNEDAAVGTSTVYTA 634	
Db	712 HYFPGVEARDHSPPLASASVTVLVNDNRPFTTKERYTLRLNEDAAVGTSTVYTA 771	
Qy	635 VRDASHVITVQITSGNTRNRPSTISQSGGLVSLATPLDYLERQYLAVTASDGTROD 694	
Db	772 VDRDANSALSYOITGANTRNRPALSTQSGMGLVTLALPLDIKOEYFPLVLTASDRALHD 831	
Qy	695 TAOIVNVNTDANTHRPVSQSHYTVNVEDRPAGTTVVLISATDEDTGENARITYFMEDS 754	
Db	832 HCYVHINITDANTHRPVSQSHYTVNVEDRPAGTTVVLISATDEDTGENARITYFMEDS 891	
Qy	755 IPQRIIDADTGAVTQAELDYEDQVSYTLATTAAPNGIPQSDPTTYELIVNDVNDNAPQ 814	
Db	892 LPQRIIDADSGAITLQAELDYEDQVSYTLATTAAPNGIPQSDPTTYELIVNDVNDNAPQ 951	
Qy	815 PLRDSYQSVYEDVPPFTSVYQISATDRSGANGVFYTPQSGDDGDDPTVESTSGIVR 874	
Db	952 PVASHRYTGVSEDAFPFTSVYQISATDRDAHANGVQTFPQNGRGGDDPTIETPSGIVR 1011	
Qy	875 TLRLIRLDENVAQYVLRAYAVDKAMPAPATPMEVTVTVLVLDVNDNPPVEQDEDFVFEENS 934	
Db	1012 TVRLIRLDENVAQYVLRAYAVDKAMPAPATPMEVTVTVLVLDVNDNPPVEQDEDFVFEENS 1071	
Qy	935 PIGLAVAVTATDPBEGNNAIMQIYVSGNIPEYQGLDIFSGELTALVDLIDREDEPEYVL 994	
Db	1072 IVGSVVAQITVADPPDGNNAIMQIYVSGNIPERFQMDIFSGELTALVDLIDREDEPEYVL 1131	
Qy	995 VIQATSAFLVRAVTVHVLIDRNDNPVYLGFEELLFNNYVTRSSSPFGAIGRPAPDP 1054	
Db	1132 VVQATSAFLVRAVTVHVLVDQNDNSPYLANNFOLLFNNYVTRSSSPFGAIGRPAPDP 1191	
Qy	1055 DISDSLTSFRRGNELSVLINAATGELKTSRALDNNRPLEALNSVLVSDGVSYTAQCA 1114	
Db	1192 DVSDDLFFSFRGNELQLLVNQTSGELRLSRKLDNNRPLVASMVLVYTDGLHSVTAQCV 1251	
Qy	1115 LKVTITDMLTHSTLRLBDSMPERFLSPLLGLFQAVAAVTLATPPRHVVVFNQGRDTP 1174	

```

1252 LKVVITTEELANSLVLENNMOERPLSLGHEGVAVALATPEDEVFITINQDND 1311
1175 APGGHILNLSVGOPEPGGSP--PPLPSEDIQERLYNRSLLTIAISORVLPDNDIC 1232
1312 V-GGTVLANSFSLALPARGAGAGAPWSSBEIQLYRRLAALARSLLDVPDNDVC 1370
1233 LREPCENYMCVSLAFDSSAPPLASSVLPRLPHVGLRCRCPGFTGDYCEYEDLC 1292
1371 LREPCENYMCVSLAFDSSAPPLASSVLPRLPHVGLRCRCPGFTGDYCEYEDLC 1430
1293 YSRPCHGCRSREGYTCLCDGYGHECEVSASGRCTPVCNKGCTCVLLNGR 1352
1431 YSRPCHGCRSREGYTCLCDGYGHECEVSASGRCTPVCNKGCTCVLLNGR 1490
1353 CDPCSGD-PEKPYCOVTTSPRPAHSFTTFRGLRORFHTLSPATKEDGILLNGR 1411
1491 CQCPAGAGFEGPCFVAARSFPSSPVMFRLQORHTLSTLSFATYQPSGLLFTGR 1550
1412 EKEDFYALVETIORQVOLTFSAGSSTTVSPFVGVSDQGMHTVOLKTYNKP 1471
1551 EKEDFYALVETIORQVOLTFSAGSSTTVSPFVGVSDQGMHTVOLKTYNKP 1610
1472 QGSEBQVAVVTDGCDYVALRGSVLENYSCAOGTGGSKSLDLTGPILLGVPDL 1531
1611 QGSEBQVAVVTDGCDYVALRGSVLENYSCAOGTGGSKSLDLTGPILLGVPDL 1670
1532 PESPVARMQOFVCMENLQVDSRHIDMADFIANNGTVPGCCPAKNCVDSNTGNGT 1591
1671 PEMFVSRKOFICOMDHLIDGRVMAAFVANNGTAGQALSHCASSPCNGELCSB 1730
1592 QMDAFSECPLEGGKSCAOMANPOHGLSSLVAN-HGLSLDIPQWYLSLMTFR 1650
1731 RMGFGFCDGCVGKXDCRLTMHPYFQNGTSLMDFGNDMEVSVWYLGSLFRT 1790
1651 GYLLQAITRGRSITTLQLEBGMVLSVEGTGLQASSLRLEPGANGDWTHNAL--GA 1708
1791 GYLLQAITRGRSITTLQLEBGMVLSVEGTGLQASSLRLEPGANGDWTHNAL--GA 1850
1709 SGCPGHAII--LSFDYQOQARAGNIAPRLHGLSLNITVGI P-----GPAGG 1760
1851 GGRGHHIIMVSLDFTLLPOTYAMGSELBELKXKHLAVGSPRPSSEBPGIIV-----G 1905
1761 CLOGVAVSDTPEBVNSL-DPSHESINVEQCSLPDCSNPCPANSYCSNDYSYSC 1819
1906 CLOGVAVSDTPEBVNSL-DPSHESINVEQCSLPDCSNPCPANSYCSNDYSYSC 1963
1820 DPGYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 1879
1964 WPGYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 2023
1880 GHPYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 1939
2024 GHPYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 2083
1940 DPGYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 1999
2084 DPGYVDCNCTNVVDLPCHEQSVCTKPSAPHGYTCEPNTYGFYCEPRLIDPC 2143
2000 PKGSF-----GTAVRHODEHGRGLPRLFNCTSIITSELKGFARLQARNESGLDS 2054
2144 PKGSF-----GTAVRHODEHGRGLPRLFNCTSIITSELKGFARLQARNESGLDS 2203
2055 LALLLNATONTAGYSGSDVAVYOLATRLAHESITORGFSATODVHTENILN 2114
2204 LALLLNATONTAGYSGSDVAVYOLATRLAHESITORGFSATODVHTENILN 2263
2115 LALLLNATONTAGYSGSDVAVYOLATRLAHESITORGFSATODVHTENILN 2170
2264 LALLLNATONTAGYSGSDVAVYOLATRLAHESITORGFSATODVHTENILN 2323
2171 VVRLDK-GNFAGA-KLPRYA--LRGQRPDLFTVTLPSVRETPPVVR----- 2218

```

```

2324 IDMEQPSSSOGAHRYPRYHNLFRGODAMDHTVLLPSQSPSPSEVLPTSSNAENA 2383
2219 AGGGAQBEELARORRHPELSQGEAVASVITYRTIGLPHANDPDGRSLRVPRPII 2278
2384 TASGVSPAPL-----EPSEBPGISTVILVYALGGLLAQOAEKRGRLPOMVW 2437
2279 NTPVVISVHODEBELLPRALDKPTVQFALLTEBERIKPICVFWMHSLVSGTGWNS 2338
2438 NSPVSVAVFRGRNFRGLVSPINLFBRLDQANRSKALCYQWDPBGADQGMWTAD 2497
2339 CEVFRNESHVSCCQNHNTSPVLDVSRRE--NEELPLKTLTYVALCVTLAALLTFP 2396
2498 CEVFRNESHVSCCQNHNTSPVLDVSRRE--NEELPLKTLTYVALCVTLAALLTFP 2557
2397 PTLRLIRSNQHGRRNLTALGLAOLVFLGIGNQADLPACTYATILHPLYCTSPW 2456
2558 VLSIRLSKNVGRGHANVAAALVABELFLGIRHTRNQLCTVALLHMFSTPFW 2617
2618 LLYGGLHYRMQVPEPRNDGRAMRFTYHALGMCVPVILGLAVGLDPBGYGNDFCM 2677
2517 DTLISFAPVAFVAVMSVFLYTLAARASCAOQGFKEGKPVSGLOSPFVALLSATW 2576
2678 LLYGGLHYRMQVPEPRNDGRAMRFTYHALGMCVPVILGLAVGLDPBGYGNDFCM 2737
2577 LLYGGLHYRMQVPEPRNDGRAMRFTYHALGMCVPVILGLAVGLDPBGYGNDFCM 2635
2738 LLYGGLHYRMQVPEPRNDGRAMRFTYHALGMCVPVILGLAVGLDPBGYGNDFCM 2797
2636 TKSSTLTSYNCPSPYADRLYO-PYGDAGSLHSTRSKSP-----SYI--PFLR 2685
2798 APGSGAVNNMTALPESGLRITLGAISTVSSAFSGRADQDOSGRSSTYLRNV 2857
2686 E-----ESALNPGQPPGLG-----DPGSLFLEGDDQCHDPTDSDSLSD 2732
2858 HSTAHABHSLQAHAGPTDLVDVAMFHDAG-----DSDSDSLSEBERL 2905
2733 SYASTHSDSEBEREEBEAFAPEQGWDSLGLGAEPLPLHSTPKCG-----PGG 2786
2906 GTPSESEBNGTRRFRPPLRAAO-----SRLLAHPDVNDNDLSTWPLAG 2955
2787 K---AP-----WPGD--FGTA--KESGNGAPEERLRENGALSHREGSLPLGSSA 2835
2956 ECEAPCALQAMGSRRLGLDSKNDKANNNDLAL--TSGD-----ETSL--GA 3006
2836 KGLKCKCLPTSEKSLRLP-----LECTSSSKGSA-----SEGRG 2876
3007 KGLKCKCLPTSEKSLRLP-----LECTSSSKGSA-----SEGRG 3055
2877 GPPRPPRQSLQOLN 2893
3056 GSLSQPASRYSRBQD 3072

```

RESULT 8  
 CLAS3 MOUSE  
 ID CLAS3 MOUSE STANDARD; PRT: 3301 AA.  
 AC 091210; Q9ESD0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cadherin BGF LAG seven-pass G-type receptor 3 precursor.  
 GN CELSR3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBT\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=2183955; PubMed=11850187;  
 RA Tisbet F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;





FT DISULFID 1974 1985 BY SIMILARITY.  
 FT DISULFID 1979 1997 BY SIMILARITY.  
 FT DISULFID 1999 2008 BY SIMILARITY.  
 FT DISULFID 2016 2029 BY SIMILARITY.  
 FT DISULFID 2031 2041 BY SIMILARITY.  
 FT DISULFID 2048 2063 BY SIMILARITY.  
 FT DISULFID 2050 2066 BY SIMILARITY.  
 FT DISULFID 2068 2078 BY SIMILARITY.  
 FT MOD\_RES 1952 1952 BY SIMILARITY.  
 FT CARBOHYD 623 623 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 2713 2713 L -> LR (IN REF. 2).  
 FT CONFLICT 3024 3024 R -> P (IN REF. 2).  
 SQ SEQUENCE 3501 AA; 358455 MW; A6B18F2DF7F4DEB6 CRC64;

Query Match 49.7%; Score 7732.5; DB 1; Length 3301;  
 Best Local Similarity 53.5%; Pred. No. 0;  
 Matches 152; Conservative 411; Mismatches 734; Indels 181; Gaps 44;

QY 157 RAGERSEESLG-----GRKKNVNTAPQFPQPSYQATVPENOPAGTFVASTIRA 205  
 DB RRGRRPGRPFPGPEAKQILSTNQARPRRAARHPQFPQYNYQTLVPENBAAGTSVLKVA 340  
 QY 206 IDPEGAGLELYMDLAFDSRNOFSLDPVTCVAVTAEBLDETSVTRTADHG 265  
 DB QDPDPGAGRLITSLALMNSRSLPSIDPQSLITPAALDESERYLRTAQDHG 400  
 QY 266 MPRRSALATLITVTNDNDPVEEQEYKESLENTVEGVLTVAATDGDAPNANIL 325  
 DB SPRLSATVAVYADNDHAPVEQAOYRTELRENTBEQPTILOAATGADAPNANLR 460  
 QY 401 SPRLSATVAVYADNDHAPVEQAOYRTELRENTBEQPTILOAATGADAPNANLR 460  
 QY 326 YRLLEGSGSP-----SEVFEIDPRSGVIRTRGVPVREVEVSQTLTVEASDQGRDGP 379  
 DB YRFV---GSPAVRTAAALAAFEIDPRSGLISTSGRVREHMESEYELVEASDQGRDGP 516  
 QY 380 STTAAVFLSVEDNDNAPOSEKRYVQVREDYTGAPVLRVTAASDQKSNVAKSIM 439  
 DB SATVRHITVLDENDNAPOSEKRYVQVREDYTGAPVLRVTAATDKDANGLVHYNII 576  
 QY 440 SGNARGFYLDACOTGALDVVSPLDYETTKETLTVRAODGGRPLSVSGLVTVQVADIN 499  
 DB SGNSRGHFAIDSLTGBIOWAPLDFEABRYALRIKQADGRPLSANTGLASIQVVDIN 636  
 QY 500 DNAPIFVSTPPOATVESVPLGIYLVHQAIDADAGNARLEKRLACVGHDPFTINGGT 559  
 DB DHAPIFVSTPFOVSLENAPLGSHVHIQAVDADHGENSRLEYSLTVAADTDFVINSAT 696  
 QY 560 GWISVAELREBYDPYSGVEARDHCTPALITASASVTVVDVNKNPFTFOPEYTVL 619  
 DB GWSVSGPLRRESEYHFPFGEARDHGSPLISASASTVTVLDVNDMPFTTKEXHRL 756  
 QY 620 NEDAAVGVSVVAVSANDRDASHVTVQITSGNTRNSPITSOGGGLVSLALPLDYLER 679  
 DB NEDAAVGVSVVAVSANDRDASHVTVQITSGNTRNSPITSOGGGLVSLALPLDYLER 679  
 QY 680 QYLAIVTASGTRQDTAQIVVNTDANTHRPVFQSHYTVNVNEDRPAQTIVVLISATDE 739  
 DB YFXLVLTASDRALHDHCYVHINITDANTHRPVQSAHYSVSNMEDRPAQSVIVVIVSASD 876  
 QY 740 DTGENARITYFMEDSIPOFRIDADTGAVTQAELDYEDQVSYTLATITARNGLPOKSDPT 799

DB DVGENARITYLLENDLEPOFRIDADSGAITLAELDYEDQVYTLATITARNGLPOKADTT 936  
 QY 800 YLELVVDVNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLNGVFTYFOGDD 859  
 DB YLELVVDVNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLNGVFTYFOGDD 859  
 QY 937 YVEVVDVNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLNGVFTYFOGDD 996  
 DB YVEVVDVNDNAPOPLRDSYQSVYEDVPFTSVQISTADPSGLNGVFTYFOGDD 996  
 QY 860 GDDGFIESTSGIVRTLRRLRENVAVQVLAAYADKMPARTPMEVTVTVLDVNPNP 919  
 DB GDDGFIESTSGIVRTLRRLRENVAVQVLAAYADKMPARTPMEVTVTVLDVNPNP 919  
 QY 997 GDDGFIESTSGIVRTLRRLRENVAVQVLAAYADKMPARTPMEVTVTVLDVNPNP 1056  
 DB GDDGFIESTSGIVRTLRRLRENVAVQVLAAYADKMPARTPMEVTVTVLDVNPNP 1056  
 QY 920 VFEODEFVVEENSPIGLAVARVATDPEGTNAQIMYOIVEGNIPVEFOLDIFSGELT 979  
 DB VFEODEFVVEENSPIGLAVARVATDPEGTNAQIMYOIVEGNIPVEFOLDIFSGELT 979  
 QY 1057 VFEODEFVVEENSPIGLAVARVATDPEGTNAQIMYOIVEGNIPVEFOLDIFSGELT 1116  
 DB VFEODEFVVEENSPIGLAVARVATDPEGTNAQIMYOIVEGNIPVEFOLDIFSGELT 1116  
 QY 980 ALVDLDEDEPEVLYVQATSAPLVSATVAVLVRNDPRLTGNELTNNYVTRSS 1039  
 DB ALVDLDEDEPEVLYVQATSAPLVSATVAVLVRNDPRLTGNELTNNYVTRSS 1039  
 QY 1117 ALVDLDEDEPEVLYVQATSAPLVSATVAVLVRNDPRLTGNELTNNYVTRSS 1176  
 DB ALVDLDEDEPEVLYVQATSAPLVSATVAVLVRNDPRLTGNELTNNYVTRSS 1176  
 QY 1040 SPFGAIGRVAHPDIDSLTYSFERGNETSLVLAATGELKSRALDNNRPLEAAMS 1099  
 DB SPFGAIGRVAHPDIDSLTYSFERGNETSLVLAATGELKSRALDNNRPLEAAMS 1099  
 QY 1177 TFGSGIIGRIPADVPVSOHLFYSFERGNETSLVLAATGELKSRALDNNRPLEAAMS 1236  
 DB TFGSGIIGRIPADVPVSOHLFYSFERGNETSLVLAATGELKSRALDNNRPLEAAMS 1236  
 QY 1100 VLVSDGVSVTAQCALRVITITDEMLTSLTLEDMSPERFLSPLIGLQVAATLAT 1159  
 DB VLVSDGVSVTAQCALRVITITDEMLTSLTLEDMSPERFLSPLIGLQVAATLAT 1159  
 QY 1237 VLVSDGVSVTAQCALRVITITDEMLTSLTLEDMSPERFLSPLIGLQVAATLAT 1296  
 DB VLVSDGVSVTAQCALRVITITDEMLTSLTLEDMSPERFLSPLIGLQVAATLAT 1296  
 QY 1160 PPDHVVVFNQRTDAPGHIILNLSVQPPRPGGP--PLRSELDQRLVANSILT 1217  
 DB PPDHVVVFNQRTDAPGHIILNLSVQPPRPGGP--PLRSELDQRLVANSILT 1217  
 QY 1297 PPDHVVVFNQRTDAPGHIILNLSVQPPRPGGP--PLRSELDQRLVANSILT 1355  
 DB PPDHVVVFNQRTDAPGHIILNLSVQPPRPGGP--PLRSELDQRLVANSILT 1355  
 QY 1218 AISAQVLPEDDNIICUREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHVGELRCRC 1277  
 DB AISAQVLPEDDNIICUREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHVGELRCRC 1277  
 QY 1356 ARSLDLVLPEDDNIICUREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHVGELRCRC 1415  
 DB ARSLDLVLPEDDNIICUREPCENTMRCVSVLRFPSSAPFLASSVLPFRPIHVGELRCRC 1415  
 QY 1278 PGFTGYCTEVDLCTSRPCGPHGCRSRBGQITCLCRDGTGEHCEVASARCRTPGVC 1337  
 DB PGFTGYCTEVDLCTSRPCGPHGCRSRBGQITCLCRDGTGEHCEVASARCRTPGVC 1337  
 QY 1416 PGFTGYCTEVDLCTSRPCGPHGCRSRBGQITCLCRDGTGEHCEVASARCRTPGVC 1473  
 DB PGFTGYCTEVDLCTSRPCGPHGCRSRBGQITCLCRDGTGEHCEVASARCRTPGVC 1473  
 QY 1338 KNGCTCVNLVGGFKDCPSGD--FEKPYCOVTRSPRSPAFSTFRGLRORFHTLLSFA 1396  
 DB KNGCTCVNLVGGFKDCPSGD--FEKPYCOVTRSPRSPAFSTFRGLRORFHTLLSFA 1396  
 QY 1474 RNGCTCVNLVGGFKDCPSGD--FEKPYCOVTRSPRSPAFSTFRGLRORFHTLLSFA 1533  
 DB RNGCTCVNLVGGFKDCPSGD--FEKPYCOVTRSPRSPAFSTFRGLRORFHTLLSFA 1533  
 QY 1397 TKERDGLLYNGRFRNKDHFVLEVIQOVOLTFNAGSTTVSPVPGSVSDGMHTVQ 1456  
 DB TKERDGLLYNGRFRNKDHFVLEVIQOVOLTFNAGSTTVSPVPGSVSDGMHTVQ 1456  
 QY 1534 TVQPSGLFLYNGRFRNKDHFVLEVIQOVOLTFNAGSTTVSPVPGSVSDGMHTVQ 1593  
 DB TVQPSGLFLYNGRFRNKDHFVLEVIQOVOLTFNAGSTTVSPVPGSVSDGMHTVQ 1593  
 QY 1457 LKTYNKPGLQGTGLPGSEBOKYAVVTVDCDTGVALRFGSVLGNYSCAAQGTGSSKXS 1516  
 DB LKTYNKPGLQGTGLPGSEBOKYAVVTVDCDTGVALRFGSVLGNYSCAAQGTGSSKXS 1516  
 QY 1594 LRYNKPRTDALGGAQGPSKDYAVLVVDQVAVVALQFAGBIGNYSCAAQGTGSSKXS 1653  
 DB LRYNKPRTDALGGAQGPSKDYAVLVVDQVAVVALQFAGBIGNYSCAAQGTGSSKXS 1653  
 QY 1517 LDTGFLILGGVDDLPESPVMRBOVFQGMRLQVDSRHILDMADFLANNGTVPCCAXKN 1576  
 DB LDTGFLILGGVDDLPESPVMRBOVFQGMRLQVDSRHILDMADFLANNGTVPCCAXKN 1576  
 QY 1654 LDTGFLILGGVDDLPESPVMRBOVFQGMRLQVDSRHILDMADFLANNGTVPCCAXKN 1713  
 DB LDTGFLILGGVDDLPESPVMRBOVFQGMRLQVDSRHILDMADFLANNGTVPCCAXKN 1713  
 QY 1577 VCDSTNCHNGGTGVNQMADFSCPCPLGFGKSCAOMANPOHFLGSSLVAM--HGLSLPTS 1635  
 DB VCDSTNCHNGGTGVNQMADFSCPCPLGFGKSCAOMANPOHFLGSSLVAM--HGLSLPTS 1635  
 QY 1714 FCSAGPCCKNNGFCBSRWGFGSCDPCVFGGTOCRILMAHPYHFGNGTSLMPCGNDMAYS 1773  
 DB FCSAGPCCKNNGFCBSRWGFGSCDPCVFGGTOCRILMAHPYHFGNGTSLMPCGNDMAYS 1773  
 QY 1636 QPWTLSIMFRTOADGVLLQA-----ITGRSTITTLQLEBGHVLVSEGTGLQA 1684  
 DB QPWTLSIMFRTOADGVLLQA-----ITGRSTITTLQLEBGHVLVSEGTGLQA 1684  
 QY 1774 VPMYLGSLRRTATKIIIMQVQVGLGPHSVLLCTGLDGLSLVTLNRAAGHTV----- 1823  
 DB VPMYLGSLRRTATKIIIMQVQVGLGPHSVLLCTGLDGLSLVTLNRAAGHTV----- 1823  
 QY 1685 SSIKLEPGRANGDMWHAQAL--GASGGPAGAI--LSPDYQOQARBAQGRILHGLHS 1740  
 DB SSIKLEPGRANGDMWHAQAL--GASGGPAGAI--LSPDYQOQARBAQGRILHGLHS 1740  
 QY 1824 -HLLDQMTVSVGRKHDLRLLEQEBGRGKHITPVSJDLFTLPQDTMAMGELQGLAKK 1882  
 DB -HLLDQMTVSVGRKHDLRLLEQEBGRGKHITPVSJDLFTLPQDTMAMGELQGLAKK 1882  
 QY 1741 NITVGGIPGPAGVA--RGFRGLQGVVSDTBEVNS--LDPHSGESINVEQSCSLPDCD 1798  
 DB NITVGGIPGPAGVA--RGFRGLQGVVSDTBEVNS--LDPHSGESINVEQSCSLPDCD 1798  
 QY 1883 QLVHGGIPSSKEEGHGLVGCIGWVIGFTFGSSALLPSSH--RVNVEPGCTVTNPCA 1940  
 DB QLVHGGIPSSKEEGHGLVGCIGWVIGFTFGSSALLPSSH--RVNVEPGCTVTNPCA 1940  
 QY 1799 SNPCPANSYCSNDMSYSCSDPGTYGDNCTVCDLNPCHDSVCTKRSASHGTTCEP 1858  
 DB SNPCPANSYCSNDMSYSCSDPGTYGDNCTVCDLNPCHDSVCTKRSASHGTTCEP 1858

Db 1941 SGPCPBHADCKDLMWTFSCRCRPGYGGCVADACLINPCQNGSCRLQGAHPGYTCDCV 2000  
 Qy 1859 PNVLAPYCEETRIDPCPRGMWGHPTGFCNCNDVSKGPPDNCNKTSGECHKENHPRPGS 1918  
 Db 2001 SGYFGQHCHENHVDQCCPRGMWSPFCGFCNCNDVSKGPPDNCNKTSGECHKENHPRPGS 2060  
 Qy 1919 PTCCLDCYCPGSLSRVCDPEDGQCPCKPVIIGRQCDRCNDPFAEVTNGCEVNYDSCPR 1978  
 Db 2061 DSCLPDCDCYPVSTSRSCAPHSQGCPCPRFALGRQCNCSDBFAEVTASGCRVLYADCFK 2120  
 Qy 1979 ALEAGIMWPRTRFGLPAAAPCKSGPRGAVRHCDBHRCWMLPNLFCNCTSTFSELKGAFAE 2038  
 Db 2121 SLRSGVWMPQTKRGVATVPCRGALGAVALRCBDQWMLBPDJFNCSTSPAFRELSLDD 2180  
 Qy 2039 RLQRNSESGLDSRSCQQLALLRLNATQTAGYFGSDVKAVALATRLAHESRQFGLSA 2098  
 Db 2181 GLEINKMLDVTYEAQKLAQRLREVTGCDHDFSQDVRTAULLAVILAFESHQCGFGTA 2240  
 Qy 2099 TQDVHFTENLLRVSGALLDPTANKRWELI-QQTBG--GTAWLQHYAVASALAAQNRH 2154  
 Db 2241 TQDAHFENELMWAGSALLAPETGHLMAALQGRAPGSGSAGLVQHLREYATLARNMEL 2300  
 Qy 2155 TYSLFTIVTPNIVSVVRID-KGNFAGA-KLPRVEA--LAGEOPDLETTVIIPESVYR 2210  
 Db 2301 TYLNVGLVTPVIMLSIDRMEHPSSTQGAARRPRYHSLMFGQDAMDPTHTVILLPSQASQ 2360  
 Qy 2211 ETPPVVRPAGPG--EAOPEPELARRQRHPELSQGEAAVAVIYRTLAGLLPHNYDPDKR 2268  
 Db 2361 PPSFVVLPTSSNAENATASVSPAPLEPSEBPISIVILLVYALGGLLPAGQARR 2420  
 Qy 2269 SLRVPKRIINTPVVISVHDBELLPRALDKPTVQVQRLIETERTPCYPMNHSILV 2338  
 Db 2421 GARLPONVWNSPVVSVAVFGRNPLRCVLVSPIMLEFRLLOTARSKALICVQMDPSPGT 2480  
 Qy 2329 SGTGMSARGEVVPRNESHVSCCQNMHTSPAVLMDVRRR--NGEILPLKTLFVVALGV 2386  
 Db 2481 DQHGKMTARDEBELVRNCSHARCSCRGTGVLMDASPRELBEDLAVFTHAVAV 2540  
 Qy 2387 TLAAALLTFPFLTLRLIRSNQHGIRNMLTAALGLAODVFLGGINQADLPRACTVIALTL 2446  
 Db 2541 SVTALVLAVALILSRSLKSNVRGIHANVAAALGVABELFLIGIRHTNQOLCTAAVAILL 2600  
 Qy 2447 HPLVICTSSWALLENLHYRALTERDVNTGPMRTYMLGNGVAPFTGLAVGLDPBXYG 2506  
 Db 2601 HYFLSTFAMLLVQGLHYRMQVPRNDVGRMRYHLAGVGPVAILGLVGLDPBXYG 2660  
 Qy 2507 NPDFCMLIYDPLIWSFAPVFAVMSVFLYIILARASCA-AQOGPEKKGPPVSGLOPS 2565  
 Db 2661 NPDRCWISIRHEPLISFAGPIYLVVWNGTWFLLAARTSCSTGQREA--KTSVILTRSS 2718  
 Qy 2566 FAVILLLSATWLLALLSVNSDPLLFLHYLPATCNCIQGPPIFLSYVVLSEVKALKLAC- 2624  
 Db 2719 FLILLVBSAWLFGLLAVNHSILAFHYLAGLGLGLAVILLFVLANADARAATPACL 2778  
 Qy 2625 SRKPSPDALTTKTLTSSYNCPSPRYADGRLYQ-PYGSASGISHSTSGSAGSOP----- 2677  
 Db 2779 GKKAAPBETRAPAPGPGSAYNTALFESSGILRITLGLASTVSSVSSASGASGAQDQDSGRG 2838  
 Qy 2678 -SYI--PLILRE-----ESALNPGQGPPIG-----DPQSLFJBGQDQDHPDPTSD 2721  
 Db 2839 RSYLRDNLVLRHGSTAHTERSLOHAGPTLDVDMFMRDACA-----DSQSD 2886  
 Qy 2722 SDSLIEDDQSGSVASTHSSDSESEEEEEEEBAAPFGQGWDSLLGPAERLPLHSTPDXG 2781  
 Db 2887 SDSLSEBRSLISPSSESEDNGRTGRFORPLRAAQ-----SERLLAHKQVDG 2936  
 Qy 2782 G-----GPGK--AP-----WPGD--FGTTA--BESSGNAPREERLENDALSRBSL 2824  
 Db 2937 NDLSIYWPALGECRAPCALQMGSEBRRLGDSNDQANNNQPELAT--TSQD---ERSL 2991  
 Qy 2825 GPLPSSAQPHKIGLKCLPTISEKSLRLPLEQCTGSSRGS-----SASRGSRG 2877  
 Db 2992 ----GAGRQRGRIGILKNR-----LQYPL--VPGSRGTRELSWCRAATIGHAV 3033

Qy 2878 PP-----PRPPQSLQEOLN 2893  
 Db 3034 PAASYGRIVAGGCGSLGSPASRYSSREQD 3064  
 RESULT 9  
 STAN DROME  
 ID STAN DROME STANDARD; PRT; 3579 AA.  
 AC 09V5N8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE protocadherin-like wing polarity protein stan precursor (Starry night  
 protein) (Flamingo protein).  
 GN STAN OR FMI OR CG11895.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL  
 RP STAGE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=20025940; Pubmed=10556066;  
 RA Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J.,  
 RA Adler P.N., Park W.J.;  
 RT "The Drosophila tissue polarity gene starry night encodes a member of  
 RT the protocadherin family.";  
 RL Development 126:5421-5429(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=99418630; Pubmed=10490098;  
 RA Ueki T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,  
 RA Takeichi M., Uemura T.;  
 RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell  
 RT polarity under the control of Fz/Dishevelled.";  
 RL Cell 98:585-595(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; Pubmed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pianthong C., Baldwin D.,  
 RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Bokoyva D., Botchan M.R., Bouck J., Brokslein P., Brothier P.,  
 RA Butts J.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Wray A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Idegawa C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kienison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,  
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupel M.P., Smith T.,  
 RA Spier E., Spadling A.C., Stapleton M., Strong R., Sun B.,

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinlock G.M., Weisenbach J.,  
 RA Williams S.M., Woodard T., Wocley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster." *Science* 287:2185-2195 (2000).  
 RU  
 RN  
 RP REVISIONS.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu I., Berman B.P.,  
 RA Bellen Court B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.,  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review." *Genome Biol.* 3:RESEARCH0083.22(2002).  
 CC -1- FUNCTION: Involved in the fz signaling pathway that controls wing  
 CC tissue polarity. Also mediates homophilic cell adhesion. May play  
 CC a role in initiating prehair morphogenesis. May play a critical  
 CC role in tissue polarity and in formation of normal dendrite  
 CC fields.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: In the pupal wing, expressed at relatively  
 CC even levels in all regions. Abundant in 6-9 hour embryos.  
 CC Expressed at higher levels in pupae than larvae.  
 CC -1- DEVELOPMENTAL STAGE: At 12 hours after puparium formation (apf),  
 CC expressed evenly at cell boundaries. By 30 hours apf, expression  
 CC is concentrated at proximal and distal cell boundaries with little  
 CC or no expression at anterior and posterior boundaries. When  
 CC prehairst emerge at 30-36 hours apf, expression becomes evenly  
 CC distributed again along the whole cell boundary.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: Contains 8 cadherin domains.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 1 GPS domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: AF172329; AAF02618.1; -;  
 DR EMBL: AB028498; BAA04069.1; -;  
 DR EMBL: AE003828; AAF58763.3; -;  
 DR HSSP: P08709; 1BP9.  
 DR FLYBASE: FBgn0024836; stan.  
 DR GO: GO:0016021; C:integral to membrane; NMS.  
 DR GO: GO:0008014; P:cellular-dependent cell adhesion molecule ac. .; IMP.  
 DR GO: GO:0005057; F:receptor signaling protein activity; IMP.  
 DR GO: GO:0016358; P:dendrite morphogenesis; IMP.  
 DR GO: GO:0007222; P:fizzled receptor signaling pathway; IMP.  
 DR GO: GO:0007367; P:segment polarity determination; IMP.  
 DR InterPro: IPRO002126; Cadherin.  
 DR InterPro: IPRO00742; EGF 2.  
 DR InterPro: IPRO01881; EGF Ca.  
 DR InterPro: IPRO06209; EGF-like.  
 DR InterPro: IPRO06832; GPCR-like.  
 DR InterPro: IPRO01879; hormn\_receptor.  
 DR InterPro: IPRO02049; laminin\_EGF.  
 DR InterPro: IPRO01791; laminin\_G.  
 DR InterPro: IPRO00203; PKD\_cys\_rich.  
 DR Pfam: PF00002; 7tm\_2; 1.  
 DR Pfam: PF00028; cadherin; 8.

DR Pfam: PF00008; EGF; 3.  
 DR Pfam: PF01825; GPS; 1.  
 DR Pfam: PF02793; HRM; 1.  
 DR Pfam: PF00053; laminin\_EGF; 1.  
 DR Pfam: PF00054; laminin\_G; 1.  
 DR PRINTS: PR00205; CADHERIN.  
 DR PRINTS: PR00011; EGF-LAMININ.  
 DR PRINTS: PR00249; GPCRSECRETIN.  
 DR SMART: SM00112; CA; 8.  
 DR SMART: SM00179; EGF CA; 1.  
 DR SMART: SM00303; GPS; 1.  
 DR SMART: SM00282; Lamg; 2.  
 DR PROSITE: PS00232; CADHERIN\_1; 6.  
 DR PROSITE: PS00268; CADHERIN\_2; 8.  
 DR PROSITE: PS00022; EGF\_1; 4.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS00649; G-PROTEIN RECEPTOR\_F2\_1; FALSE\_NEG.  
 DR PROSITE: PS00650; G-PROTEIN RECEPTOR\_F2\_2; FALSE\_NEG.  
 DR PROSITE: PS00227; G-PROTEIN RECEPTOR\_F2\_3; 1.  
 DR PROSITE: PS00261; G-PROTEIN RECEPTOR\_F2\_4; 1.  
 DR PROSITE: PS00025; LAM G-DOMAIN; 2.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 1.  
 DR PROSITE: PS00221; GPS; 1.  
 KW Cell adhesion; Developmental protein; G-protein coupled receptor;  
 KW Calcium-binding; Repeat; Signal; Transmembrane; EGF-like domain;  
 KW Laminin EGF-like domain; Glycoprotein.  
 FT CHAIN 1 29  
 FT 30 3579  
 FT  
 FT DOMAIN 30 2816  
 FT TRANSMEM 2817 2837  
 FT DOMAIN 2838 2845  
 FT TRANSMEM 2846 2866  
 FT DOMAIN 2867 2883  
 FT TRANSMEM 2884 2904  
 FT TRANSMEM 2905 2919  
 FT TRANSMEM 2920 2940  
 FT TRANSMEM 2941 2959  
 FT TRANSMEM 2960 2980  
 FT TRANSMEM 2981 3000  
 FT TRANSMEM 3001 3021  
 FT TRANSMEM 3022 3031  
 FT TRANSMEM 3032 3052  
 FT TRANSMEM 3053 3579  
 FT DOMAIN 360 464  
 FT TRANSMEM 465 581  
 FT TRANSMEM 582 689  
 FT TRANSMEM 690 794  
 FT TRANSMEM 795 897  
 FT TRANSMEM 898 1007  
 FT TRANSMEM 1008 1113  
 FT TRANSMEM 1114 1220  
 FT TRANSMEM 1482 1518  
 FT TRANSMEM 1556 1753  
 FT TRANSMEM 1756 1792  
 FT TRANSMEM 1795 1963  
 FT TRANSMEM 1965 2000  
 FT TRANSMEM 2091 2126  
 FT TRANSMEM 2744 2802  
 FT TRANSMEM 140 143  
 FT TRANSMEM 140 159  
 FT TRANSMEM 155 159  
 FT TRANSMEM 2567 2579  
 Query Match 31.0%; Score 4817.5; DB 1; Length 3579;  
 Best Local Similarity 34.9%; Pred. No. 68e-223;  
 Matches 1108; Conservative 461; Mismatches 1005; Indels 599; Gaps 71;  
 QY 171 RRRNVTAPOPPPSYQATVPENQAGTPVASIRAIIDPDGEAGRLTYTMDALFDSRSNQ 230  
 Db 349 REELRNQSPYFQALYVASVLEEQPAGAAVTVRRARDPDS---VVYSVSLDLSRSQS 405  
 QY 231 FSLDPVTCATVTAELDRETSHTVPRVTADQHGKPRASALATLTILYTDNDHPDVE 290

Db 406 LFKVDSRIGVTTSSALDELMDVHFRVVAATDSEPPRSQTTTLQVNLDCNDHSPTE 465  
 Qy 291 QOEYKESIRENLEVEGYELTVRAATDGDAPPNANILY---RLTEGSGSPSE---VFBIIDP 344  
 Db 466 AEGFASIREGATVSGTVTLRATDODIGKMEIEIGIAVVDGAGLADQCMPIFRIDS 525  
 Qy 345 RSGVIRTRGPVDRREVESTQLTVEASDQGRDPGRST-TAAVFLSEVEDNDNAPQFSER 403  
 Db 526 RSGVITSSSLDRETSDSYHLVTLADLASAQSERRTASVQVKVLDNDNYPQFSERT 585  
 Qy 404 YVVOVREDVTGA---PVLRVTAASDRDKGSNAVHYHYSIMSGARQOFTLADQGLDVS 460  
 Db 586 YVVOVPEDEMGSTEDNTVAHIRATDADQGNNAIRALIGANTQSOFSIDSGVSLVK 645  
 Qy 461 PLDYETTKYELTVRAADGGRPLSNVSGLVTVQVLDINDNAPITVSTPQATVLESYL 520  
 Db 646 PLDYESVSYRLVIRAQDGSFSGRNTQQL-VNVIDANDNAPRYTQFQSVLENV 704  
 Qy 521 GYVLHVQALIDAGDNARLEYLAGVGHDPFTTNGTGWISVAELDREBVDYSGV 580  
 Db 705 GNIIIRVQAYDSDEGANMEITYSISERDNPFLAVDPRTGWQTIKPLDREOGRFAQV 764  
 Qy 581 BARHGTBALTASASVTVLDVNDNPTFQPEYTVLNDPAAGTGVWVYSAIDRAH 640  
 Db 765 VAKDGVPEPKSASSSVITVODVNDNDPAFNKYEANVEDQPCPTVTVTAADPDED 824  
 Qy 641 SVITVQITSGNTRNFSITSGSGGLVSLALPLDYKEROVYLAATASD-GTRODTAOIV 699  
 Db 825 SRLHEIITGNTRGFAITSONRGILITAGSLDQKGRPLITAAIDSGSRSTAVH 884  
 Qy 700 VAVTDANTHRPVQSGSHTVVANVEDRPAQTVVLIISATDEBGTENARITTYM-EDSI--- 755  
 Db 885 INIDANNAPRIFFENAPYASVFEBAVQGTVLVVSATSDSDGVAQITYSINESINGL 944  
 Qy 756 ---POFRIDADPQATQABLDYEDQVSYTLATARDNGIPQKSTTLELIVNDVNN 812  
 Db 945 GSPDPFSINPOTGALVTNAPLDRETTSGYLTITAKDGSNLSLITDVEIGVTDVNN 1004  
 Qy 813 POPLRDSYQSGVYEDVPPTSVLQISATDRDGLNGRVYTFQGGDDGDFIVESTSGI 872  
 Db 1005 PAFKSPFLYQASILBDAVGTSTVIVAASDPDGLNGRIKYLISDRIDEDGDFVDPSTGT 1064  
 Qy 873 VRTLRDRBENAVQVVAAYVNDKGMPPARPTMETVTVLDVNDNPPVPEDEBEPVPEE 932  
 Db 1065 IRTNGGLDRBSVAVHLTALAIDKSPPLASTVEQIIRLEBDVNSPPTFADDKITLYPE 1124  
 Qy 933 NSPGLAVARVATADPDEGTNAQIMYQIVEGNIPEVFOLDIFSG---ELTALVDLVED 988  
 Db 1125 NSPVQSVGEIHADHPDGCNAVHYHSITIGDDNSAFSLVTRPGBRAQLMTWELDYES 1184  
 Qy 989 RPEVYVLIQATSAPLVSRAVTVHLLDRNDNPVLANFEILFNNTVTRSSSPGGAIG 1047  
 Db 1185 TRKFELVIRRAAPPLRNDNAHEILVTVNDNAPVLRDFOYIFNNF---RDHFPSSIG 1240  
 Qy 1048 RVPAHPDISLSLTYSPFERGNELSVLNASTGELKSRALDNNRPLEAINSVYSDGVH 1107  
 Db 1241 RIPADADVOSKLRHILISGNANMLRLNSSGGVLVSPOLNTVNPKATVEVSQDIN 1300  
 Qy 1108 SVTAQACALRVITLEDMLTHSTITLLEBMSPERFISPLGFIQVAAVTLATPPDHVVF 1167  
 Db 1301 EAKIMQSVRLITEDMLFNSVTVLNTMTERRAFISPLNPFLOGIAALITCPKEHIVF 1360  
 Qy 1168 NVORDTPADGGHILANVLSVQOPPGGPPPLSEBDOERLYLNRSLITAIASORVLP 1227  
 Db 1361 SIODTIV-SSRIINVSFSARRP--DVSHFEYTFQYQERVYINRAILATIALEVLFP 1417  
 Qy 1228 DNICLRPCENYRCVSVLPDSSAPFIASSSVFRPHVYGIRGCPGPTGD----- 1283  
 Db 1418 DNICLRPCENYRCVSVLPDSSAPFIASSSVFRPHVYGIRGCPGPTGD----- 1283  
 Qy 1284 YCETEVDLCYSPGPHGRCSRREGYATCLCRDGYTGEHCVSARSGR-CTPGVCXKNGT 1342  
 Db 1478 LCPTEVDLCYSPGPHGRCSRREGYATCLCRDGYTGEHCVSARSGR-CTPGVCXKNGT 1342  
 Qy 1478 LCPTEVDLCYSPGPHGRCSRREGYATCLCRDGYTGEHCVSARSGR-CTPGVCXKNGT 1342  
 Db 1478 LCPTEVDLCYSPGPHGRCSRREGYATCLCRDGYTGEHCVSARSGR-CTPGVCXKNGT 1342

Qy 1343 CVNLLVGFKDCSPGDEKRY---COVTTSPFASFTFRGRLORPHFTALSPATK 1399  
 Db 1538 CLS-----NYSSQ-PPRYTATCELBARAFBRNSFLITRESLKORHFRFLKARFATVQ 1588  
 Qy 1400 RDGILLNGRPNRKHDFVALBEVIOEQVLTFSAGSESTTVSPVPGVSDGOMHTVOLKY 1459  
 Db 1589 ENGLILNGRYNEIHLDTALEIHEGHVSPFSILDHSEISVIOEAKVSDCKMHQVEVY 1648  
 Qy 1460 YNRPILQOTGLPOGSPQKAVVTVVDCDPTGVALRPFSSVLANTSCAAGTQSGSKS--- 1516  
 Db 1649 LNRSV-----TVLNDCTFALIS-GQLGDRNSCANRTTLKDKRSSL 1691  
 Qy 1517 -----LDLTGPELLGVDLPESFPVRARQFQGCNRIQVDSRHIDMADFIANGTVP 1569  
 Db 1692 TETCHRLDLTGFLQVGVGLRIPAHFVTNRDPFGCISDLRIDRFDLANSYVADNGLIA 1751  
 Qy 1570 GCPAKKNCDSNTCHNGCTCVNOMDAFSCBCLPAGSKSCAQEMANPOHFLGSSLVAMHG 1629  
 Db 1752 GCPQKAPLCOSEPCFNGCTCRBKGITVSCBCEGYAGNSCODNIPAWRFSGDLSFNP 1811  
 Qy 1630 LSLPISQPYLSMFRROADGVTLQAITRGRSTITQLREGHVMSVEGGLQASSLRL 1689  
 Db 1812 LARPLOUPWTSSLRTRQKBAFLQIQONSSAAVCLQGVLYTIFDEBPMYLAAGFL 1871  
 Qy 1690 EPGRANDGMHHAQALAGSGRGAHILSFYQGOARAGNIGPRLHLSNITVGGIPG 1749  
 Db 1872 -----SGBWHNRVIR-MQGSSEH--FSVDYQGRSGSVMSQVGLVYGIKIMSGPDG 1923  
 Qy 1750 PAGVY--ARGFRGLQGVAVSDTPEGVNSLDPHSHESI-----NVEGCSLPDPCDS 1799  
 Db 1924 SIGVPEASPEEGCTQDVR-----GAGGSVLSPRTIRENVEBGCSRAQCPD 1971  
 Qy 1800 NPCRANSGNDMDSVSCSDPGYVGNCTNVCDLNPCEHOSVCTRPSAPHGTGEC-P 1858  
 Db 1972 H-CPNHSQSSMDLSCEBDSGIVGTDCAPICTVPCA-SGVCNANTSIPRGIDCCNS 2029  
 Qy 1859 PNYLGPYCETRIDQPCERGMWGHPTGCPNCVDVSKGFPDPCNTSGECHKENHYRPPGS 1918  
 Db 2030 SSRHGDYCEKELQPCPCGMMGWRVCGPCRCDLAOGVHPCKNTTQGCYCKINHYQPPNE 2089  
 Qy 1919 PTCILDCPYTSGLSRVCBDEBQCCPKRGVITRQCDRCNPRAYVTNCCVNNYDSCR 1978  
 Db 2090 TACISDCYISIGSPSGACNPLTQCECRBGVIGRCDSNSPFAEVLTSCVEVYDSCR 2149  
 Qy 1979 AIEAGIWWPRTFRGLPAAAPCPKSGFETAVRHCDERHG-WLPNLFNCTSIETSELKGA 2037  
 Db 2150 SFGAGVWMPRTPIGVALIBCCPPRANGKGRSCDVQSGSNTEDMNCTSEPFVELARQL 2209  
 Qy 2038 ERLQNESGLDSGRSQALALLRNA-----TOHTAG----- 2068  
 Db 2210 SOLEKLELEINSFVALIKMAEQLRKACEAVDRGASKQKISGNGRPNRRYKMESSFLLSN 2269  
 Qy 2069 -----YFGSDYKAVVQALATRELLAHESQORGFASATDVHF 2104  
 Db 2270 GGNVMSHELEMDYLSBELKFTHDLRYGADLVTEGGLIQLIELINLEWQSNLHSQKYF 2229  
 Qy 2105 TENILRVGSLALDTPANKRH-----ELIQTEGEGTAMLLQHYEAVASALQONMHTYLSPP 2160  
 Db 2230 IKVLVDAASVILDRKYEABERRATELOR--GPDDLVDFAFNKYLIVLARSQDITYSP 2386  
 Qy 2161 TYTTPNIVISVRLDKNFAGATLPRYEALRGOP-----PDELT--VILP 2205  
 Db 2387 EIVQPNMAL-----GLDIVTSESIFGEPEOLSEYHRSKYLKPNAPFTTESVLP 2435  
 Qy 2206 ES-----VERNPPVVRP-----AGPGEAOPBELAR 2233  
 Db 2436 DTSGFLQHSARQRPVVSFPKYNNTIIDRRKFDQHTKVLVLEMLGITPESDEISQSGRR 2495  
 Qy 2234 QRRHPELSQGEAAVASVITYRTLAGLPHNYD----- 2264  
 Db 2496 GSSHHDH-----RAIVAAQYVDVGLPDLVDRTTRKMGVDVELANPILSLQILVPSMER 2551

QY 2265 -PDKSLRVKRPRLINT----- 2280  
 DB 2552 EOEJRLERIPERKLPSSSSSSSSSGSTEQFVEVDPVAPRTSSSEQIEDIRITAHB 2611  
 QY 2281 ---FVBSI-----SVHDEELP----- 2295  
 DB 2612 IPPVSSVEQGEASDDEGERERHRLNLDIFHNGSGEVLSPDSEPMANPYGVS 2671  
 QY 2296 -----RALDXPTVQF-----RLLETER----- 2313  
 DB 2672 STGSDQPKGENEAVYEDRLVYKQVEITYPSEQMOQEQVYVNSLSPHLAQIKLQMW 2731  
 QY 2314 -----RTKICVFNHNSILVSGGWSAGCEV-----VFNESHVSCQ 2352  
 DB 2732 LDVDSARFGRPSNQCVRWN-----SFTNQMTRLGCGCTEIPDFGDENPAQAQALIVNCS 2786  
 QY 2353 CNHNTSPAVLMDVSRRENGEILPLKLT-----TYVALGVTLLALLTFEFLTLRLRSN 2407  
 DB 2787 CTHISSYAVIVDIDPDP-----IPEPSILVQITSASAPLVSLPILGVTLLALALRGQOTN 2843  
 QY 2408 QHGIRRLTALGLAQVFLGIGNO-----ADLPACTVIALHFLYLCTPSWALIEALH 2463  
 DB 2844 SNTIHWNI VCVFCABELLFVGMQSRRLSEFPCKLTALCLHYFWLAFAWTTVDVCH 2903  
 QY 2464 LYRALTEVRDVTGNPMEFYTMGCVAFITGLAVGDDEPGNPDCTMSTYTLTWSF 2523  
 DB 2904 LYRLTETMRDINHPGPFYPMGGAIPAIVGSLVGRABEYGNLSLCTMSTYVEPVWML 2963  
 QY 2524 AGPAPAVMSVFLYIILAAASCAARQGFQFEKQPVSG-----LQSPFVALLLSATV 2576  
 DB 2964 VGPILAGSVNLLIFVSVDA-----FTLKQHVIGFGKRLTLMLSVSLPLMGVMW 3016  
 QY 2577 LIALLSVNSPTLLFHYLPATCNCIQGPFILSYVLSKEVRKALKACSR----- 2626  
 DB 3017 VLAIVLAASEHSQSLSLISGVVLLHAFCLIGCIITNKVRVENVQRTCLRMGKRVPLD 3076  
 QY 2627 -----KPS-----PDPALTKSTLTSSVNCSPFADG 2653  
 DB 3077 SSNWNSSSHNNAAPNSFNLASGYTTTRNIGISASTSTSTKTS-----SSPYSIDG 3133  
 QY 2654 RLYQPYDASGLSHSTRSGKQPSYIPLILREBSALNPGQPGGLDPSLFLGQDQ 2713  
 DB 3134 QLMQT-----STSTSNVNSASDAPSLRGSESTTGRSGEKEP-----SKROR 3178  
 QY 2714 HDPEPTDSLSLEDQSGSYASTHSSDEEBE-----EEEE 2751  
 DB 3179 KDSOSGET-----DGRSLKSLASHSSDDDESRATSGSTRSTAVSTPAVLNITEHV 3233  
 QY 2752 EAPAPGE-----QGDLSLGP-----GAERLP-----LHSTPKDGGGPGKAPMPDPR 2796  
 DB 3234 QATTPPLNVVQSPQLPSPVAKPYAPARWSSQLPDAYLQSPPING-----RMSQPTGS 3286  
 QY 2797 TAKSSNGAPBERTLRNGDALSRGSLGPLPGSSAQPHKGLKKCLPTISEKSL----- 2853  
 DB 3287 DNEHVHQA-----KMTISPNLPLN-----PULTTSTYLOQH 3318  
 QY 2854 ---LRLP---LEOCTSSSRSSASSESGRGGPP-----RPPROSLQEOQLNG 2894  
 DB 3319 HNKINMPISILNIDARBEYEDSLVGRGEYDPKYGSYMPGSHYSGEKDYPG 3371  
 RESULT 10  
 PAT DROME STANDARD; PRT; 5147 AA.  
 AC P33450; Q9V0X5;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cadherin-related tumor suppressor precursor (fat protein).  
 GN FT OR CG3352.  
 OS Drosophila melanogaster (fruit fly).  
 OC Burkholderia; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

CC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92069752; PubMed=195913;  
 RA Mahoney P.A., Weber U., Onofrechuk P., Bissmann H., Bryant P.J.,  
 RA Goodman C.S.;  
 RT "The fat tumor suppressor gene in Drosophila encodes a novel member  
 RT of the cadherin gene superfamily.";  
 RL Cell 67:853-868(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 KC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Ewanicki A.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
 RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclib J.M.,  
 RA Palazzo J.M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Sytzkies R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN. ACTS AS A  
 CC TUMOR SUPPRESSOR. REQUIRED FOR CORRECT MORPHOGENESIS.  
 CC -1- SIMILARITY: Contains 34 cadherin domains.  
 CC -1- SIMILARITY: Contains 5 EGF-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-bb.ch](mailto:license@sib-bb.ch)).  
 CC EMBL; M80537; AA28530.1; -.  
 DR EMBL; AE003577; AAF51036.1; -.  
 DR HSSP; P00740; 1EDM.  
 DR FLYBase; Fgn0001075; ft.  
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.  
 DR GO; GO:0008014; P:calcium-dependent cell adhesion molecule ac. .; NAS.  
 DR GO; GO:0015339; P:calcium-dependent cell-cell adhesion; NAS.

DR	GO: 0008283;	P:cell proliferation; IMP.
DR	GO: 0000904;	P:cellular morphogenesis during differentiation; IMP
DR	GO: 0045317;	P:equator specification; IMP.
DR	GO: 0007446;	P:magnal disc growth; IMP.
DR	GO: 0018149;	P:protein-protein cross-linking; IPI.
DR	InterPro: IPR002126;	Cadherin.
DR	InterPro: IPR000742;	EGF 2.
DR	InterPro: IPR006209;	EGF-like.
DR	InterPro: IPR001791;	EGF-like.
DR	InterPro: IPR001791;	Laminin_G.
DR	Pfam: PF00008;	Cadherin; 34.
DR	Pfam: PF00054;	laminin_G; 2.
DR	PRINTS: PR00205;	CADHERIN.
DR	SMART: SMO0112;	CA: 34.
DR	SMART: SMO0181;	EGF: 4.
DR	SMART: SMO0282;	LamG: 2.
DR	PROSITE: PS00232;	CADHERIN 1; 22.
DR	PROSITE: PS50268;	CADHERIN 2; 34.
DR	PROSITE: PS00022;	EGF 1; 4.
DR	PROSITE: PS01186;	EGF 2; 2.
DR	PROSITE: PS50025;	LAM_G.DOMAIN; 2.
KM	Cell adhesion; Signal;	Transmembrane; Glycoprotein; Calcium-binding; Repeat; EGF-like domain.
FT	CHINAL	1 35
FT	CHINAL	36 5147
FT	DOMAIN	36 4583
FT	TRANSSEM	4584 4609
FT	DOMAIN	4610 5147
FT	DOMAIN	36 156
FT	DOMAIN	157 270
FT	DOMAIN	271 382
FT	DOMAIN	383 494
FT	DOMAIN	495 599
FT	DOMAIN	600 708
FT	DOMAIN	709 820
FT	DOMAIN	821 942
FT	DOMAIN	943 1049
FT	DOMAIN	1050 1153
FT	DOMAIN	1154 1278
FT	DOMAIN	1279 1384
FT	DOMAIN	1385 1489
FT	DOMAIN	1490 1601
FT	DOMAIN	1602 1713
FT	DOMAIN	1714 1823
FT	DOMAIN	1824 1922
FT	DOMAIN	1923 2027
FT	DOMAIN	2028 2167
FT	DOMAIN	2168 2278
FT	DOMAIN	2279 2385
FT	DOMAIN	2386 2491
FT	DOMAIN	2492 2596
FT	DOMAIN	2597 2703
FT	DOMAIN	2704 2810
FT	DOMAIN	2811 2913
FT	DOMAIN	2914 3013
FT	DOMAIN	3014 3124
FT	DOMAIN	3125 3229
FT	DOMAIN	3230 3334
FT	DOMAIN	3335 3439
FT	DOMAIN	3440 3545
FT	DOMAIN	3546 3651
FT	DOMAIN	3652 3756
FT	DOMAIN	3950 4011
FT	DOMAIN	4013 4049
FT	DOMAIN	4052 4090
FT	DOMAIN	4092 4128
FT	DOMAIN	4129 4320
FT	DOMAIN	4321 4362
FT	DOMAIN	4402 4569
FT	DISULFID	3954 3966
FT	DISULFID	3960 3999
FT	DISULFID	4001 4010

FT	DISULFID	4017	4028	BY SIMILIARITY.
FT	DISULFID	4022	4037	BY SIMILIARITY.
FT	DISULFID	4039	4048	BY SIMILIARITY.
FT	DISULFID	4056	4067	BY SIMILIARITY.
FT	DISULFID	4061	4078	BY SIMILIARITY.
FT	DISULFID	4080	4089	BY SIMILIARITY.
FT	DISULFID	4096	4107	BY SIMILIARITY.
FT	DISULFID	4101	4116	BY SIMILIARITY.
FT	DISULFID	4118	4127	BY SIMILIARITY.
FT	DISULFID	4325	4341	BY SIMILIARITY.
FT	DISULFID	4334	4350	BY SIMILIARITY.
FT	DISULFID	4352	4361	BY SIMILIARITY.
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	257	257	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	402	402	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	461	461	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	605	605	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1155	1155	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1367	1367	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1458	1458	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1751	1751	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1831	1831	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1880	1880	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2080	2080	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2171	2171	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2247	2247	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2437	2437	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2581	2581	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2799	2799	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2920	2920	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2946	2946	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2967	2967	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3167	3167	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3303	3303	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3386	3386	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3389	3389	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3525	3525	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3852	3852	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3865	3865	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	3905	3905	N-LINKED (GLCNAC. . .)

Query Match	9.9%; Score 1534.5; DB 1; Length 5147;
Best Local Similarity	22.8%; Pred. No. 4e-65;
Matches 549;	Conservative 291; Mismatches 665; Indels 903; Gaps 72;
QY	176 NTAFQFPSPSYOATVVPENQDPAGTFVASTRAIDPDEGAGRLLEYMDALFDSRSNQPSLD 235
DB	2273 DNGSVFDPKQYSASVAENASIGAMVLOVSATDDEGANGKIRISI-VLGD-QNHDEGIS 23299
QY	236 PVTGAVTTAAE----- 246
DB	2330 EDTGVAVAKNKLNERLSRYSLTVRAEDCALENPAGTAEITINILINDNRPTFLDSPY 2389
QY	247 -----L 247
DB	2390 LARVMENTVPNGGYVLTVNAVDADPTPLNSQVRYFLKEGSDLFRINASSGDIALLKPL 2449
QY	248 DREKTSHTVFRVNTAQDGMPPRSALATLITLVDTNDHDPYFEGQAFKESIREULEVGYE 307
DB	2450 DREQOSEETLLTVAMDGSSPPLTGTGIVRVEVDIINDNPVFQIASHATVREMLPGGTH 2509
QY	308 VLTVRATDGAAPNANTLLRYLLBESSGGSPESEVFIDPDRSGVITFRGVDREEVSYOLT 367
DB	2510 VLTFRATDKEGELNAKIRFNL-----GEHMRFRFIDSETGISTATLLDREISVYHLLT 2565
QY	368 EASDQGDGPGRSTTAAPLAVLSEDDNDNAPQFSSEKRVVVOQREBDVPGAPLRTASDRD 427
DB	2566 MAQSSSITE-PRASVVLITISVDVNNIKFSDSTYNNAVPERISNGEYVFGARALDD 2624

QY 428 KGSNAVHYSTIMSGNARQFYTLDAOT-----GALD--- 457  
 Db 2635 DGENAVVHYTL---SGRDQHFYDINTKTGVSTKLELTKTKSHDDLTYTTLVISAMDGE 2681  
 QY 458 -----VSP----- 461  
 Db 2682 QSLSKELJYLIRPELPTPTAYMANSHFAMSEVBRCKKITTKSATSPKGLVGYKIRY 2741  
 QY 462 -----LDYETTKETTLVRADQDGRPRPLSNVSGLVYQV 495  
 Db 2742 AIAAGIMGDSLVRPNSGLLSVGQGLDYELTHLYEIMIBADGOTPSLRVTL-LITLNV 2800  
 QY 496 LDINDNAPIFVSTPQATVLESVPGYVLAHQALDADAGNARLETRLAGVGHDFPTI 555  
 Db 2801 TDADNAPVEMOLLYNAEVLBEESPQOLIAVYKASDRSDGNGVYTLQ---NDPQDTP 2857  
 QY 556 N-NGTGMISVAALDREBVDYFSGVEARDGTPALTSASVSYVLDVNNNNPFPQPE 614  
 Db 2858 BITSGELTYMRDRREIBIDYAFVVEAVDQGVHMTGTASVLLHLDKNDNPPFTR-L 2916  
 QY 615 YTRILNEDAVGTSVVTVAVDRO--AHSVITTYOITSNTNRRPSITSQGGGLVSLALP 672  
 Db 2917 FSLNVTENAEIGSFVIRVTSDDLIGANANASYSF--SENPKCFRIEPOGN--ITVAGH 2973  
 QY 673 LDYTLERQYVLAVTAASDGTRODTAQIYVNTDANTHAPVQSSHYTVNNVEDRAGTTVV 732  
 Db 2974 LDRQOQBEYILIKVAASDAMRAETPTITITDQNDNAPFESHSPYSFSPFELQOSIALVG 3033  
 QY 733 LISATDDE--TGENARITYFMEDSIPOFRIDADTG-----AVT 768  
 Db 3034 QIATDDBKQPNVSVISLQSPSPMSIDATSEVSKAVRKHQYVARSPEMMTALT 3093  
 QY 769 TOA----- 771  
 Db 3094 VLATDNKCPPLYSECLVNVINIVDAHNNPKFEQAELYALPQDAVGRQIRVHANDKOD 3153  
 QY 772 -----ELUYE-----DOVSYTLATIPARNGIPOKSDTT 799  
 Db 3154 LGTNEMOYSLMTFNLSISPSYGRHDGMITLVKPIQVPPNTRYELVVAATRGVPPQSDBT 3213  
 QY 800 YLEILVNDVNDNAPOFLRDSYQSVYEDVPFTSVLOISATDRDSGNRGRFYTPQSGDD 859  
 Db 3214 RVYLVVVGEMMDPRFVSNVQVYVPENEPVGTYLLVGAITDDDTGNGHRLRISISGANE 3273  
 QY 860 GGDGFIYESTSGIVRTLRIDRENAQVYLAHAVDKMPARTPMEVTVYLVNDNPP 919  
 Db 3274 RQ-DPFSVDERTGGIVIOQLDYDLIQEYHLNITVQDLGYHPLSSVAMLTITLTDVNDNP 3332  
 QY 920 VFPQDEEDVFEENS----- 935  
 Db 3333 VFNHKEHYCIYIPEKPVGTVPQAHAAADKOSPKNALIIHYALPSPGDPHHPFIMNGQTI 3392  
 QY 936 ----- 935  
 Db 3393 SSASVFPYEBRRIYTLQIKAKNPDSMESYANLYVHVLGNERFPQPLQVHFHFVDSETS 3452  
 QY 936 -IGLAVARVATPDDEG----- 951  
 Db 3453 AVGTFRVAVOATDKDGEDGRVYLLVGSNDKGRIDTWTGLIYVARHLDRBTQNRVVL 3512  
 QY 952 ----- 951  
 Db 3513 TYNANANYSIRGNDTDEAQVYIISIQDNDEPEFIKHYYTSTISBAFVGTKVTYVKAIDK 3572  
 QY 952 -----TNAQIMYOIVEGNIPEVFOGLDIPSGELTALVDLYEDREPVYLVIOATSA---PLV 1004  
 Db 3573 DVARTONNQFYSIISINGMLKOSFKIDVOTGISTASRLDRREBTSTYMLYGAIDTGLPQOT 3632  
 QY 1005 SRATVAVRLLDKNDNPPVL---GNFELLNNYVYTNSSSPGGAIGRVPRAHPDI---SD 1058  
 Db 3633 GSATVHIELEDDVNDNGPFTPEG---LNGYISENBP--GTSIMTLIASDPLPRNGG 3685  
 QY 1059 SLTYSFPERGHELSVILNASTG----- 1080

Db 3686 PFTYOLIGKHKSWLSDVRNSGVVRSYTSFDEBMTPIIEALIEVEDSGKPKQSGHLLTI 3745  
 QY 1081 -----E 1081  
 Db 3746 TVLDQNDNPTSTRSLHIAVSLFNGDLPSNVKLAADVBRNDIDIVGDYRCRLQKNAPSOLO 3805  
 QY 1082 LKLSRALD-----NRPLEAIMSVLVSDBGVH--SVTAOCALRVTLITDEMLTHSITLRLED 1135  
 Db 3806 LAIPRACDLITTSHTTPIASVPSYTGNDGKHGDVSSKVSVAFQSPNNETLANSVISWVRN 3865  
 QY 1136 MSPERFLSPILGLETOAVALATATPPDHVVVFNVOQRTDAPGGHILNVSL-----S 1186  
 Db 3866 MTAYHFLANNHYRPILEMIKSRMSN--EDREVILYSI--LEGSGNSTVLOLMAVRLAKTS 3921  
 QY 1187 VGOBPGGGGPPPLPSDLOERLYLNSLLTALSACVLPFPDNICLRECEMTRCVSV 1246  
 Db 3922 YQDP-----KTLERLERKRSAPSELOKREVIVGYEPCSPDVCEMNGVCSAT 3969  
 QY 1247 LRPDSAPFLASSS---VLFRPIHPVGLRCRCPPTGD----- 1283  
 Db 3970 MRLIDHNSFYIOPSPLVLSGP--RVNHVDSQCCTSGSGBQCSRRQDPCLPNCHSOVOC 4028  
 QY 1284 -----YCTE--VDLCYSRPGCPHGR--RSBEG--GYTCLCBGYTGEH 1322  
 Db 4029 RLUGSPQCMCPANRDKHCEKERSDVCYSKPCRNQSGCORSPDSSYFCLCRGFRGNQ 4088  
 QY 1323 CEVSARSGRCTPGVCNNGGTCVNLVGVKFCDCPSGDFEKPQYQVTRSPHNSFTFRG 1382  
 Db 4089 CE--SVSDSCRPNCPCHGGCLVS--LKPQYKCNCTPGAYG--HCBREYFGQPLSYMTFPA 4144  
 QY 1383 LRORFHTLASFATKERDGLLYNGRFNE--KHDFVLALEVOVOLTFSAGESTTVS 1440  
 Db 4145 LDVTYN--DISIVPATIKPNSLILYNGMGSGSDPLAIELVHGRAY--FSSGARPAIS 4201  
 QY 1441 PVPFG--GVSDQGHVYVQLKTKNPKLLQGTGLFGPGBQKVAVTVVDCGTVALRFGSVL 1499  
 Db 4202 TVIAGNLDGGMKV-----TATRNG-----RWSLSVAKCADSGVCTECLP 4245  
 QY 1500 GNYSCAAQ-----GTQGSKKSLDTGPLLIGV-----PDLPSFPYRMRQFVGMRLQ 1550  
 Db 4246 GDSSTADREVGPVGTIANFNQ-----PLMTIGLSADPLFERPGOVHSDDLVGLHSVH 4299  
 QY 1551 VDSRHIDMDFIANNQTVPGCPAKKNVCD---SNTCHN--GTCVNAQMDAFSGCEC----- 1600  
 Db 4300 IGRALNLSLPLQOKGILAGC--NRQAQCPALAAHRCGGAQCIDRMSSSLQCCGHILQ 4357  
 QY 1601 -----PLGGGKSCAQ---EWANPOHFLG---SSLVAV-----HG 1629  
 Db 4358 SPDQSDLSLEPITLIGBAPFVEFRISBYRMOQLLDNLNYSKSAVLDNQWRERRAVGNFT 4417  
 QY 1630 LSLPISQPWYLSLMPRTQADGVLLQATIRGSTILOLREBHV-----LSVBGTGLQ 1683  
 Db 4418 ASQIYAPKMLSLFTFYQOQGITLYAATNQMT--SLSLBGLVYYSKHOLITNTVOE 4476  
 QY 1684 ASLSLEPGRANDGMHHAQLALGASGPGHAILSFYQOQRA-----EG-NLGPRLH-- 1735  
 Db 4477 TSTL-----NDGKHNVS-----FESNSLRILVGRQVGDLDLA 4513  
 QY 1736 GLH-----LSNITVGGITPGAGVARGRCIQGVSDTPEGVN---SLDPS---HBE 1783  
 Db 4514 GYVHDFLDPYLTILNVG-----EAFVGLCANVTANNELOPLNGSGSIFPEVAYHGK 4564  
 QY 1784 SINVEQGC 1791  
 Db 4565 ---IESGC 4569

RESULT 11  
 PATH HUMAN  
 ID PATH HUMAN  
 AC 014517;  
 DT 16-OCT-2001 (Rel. 40, Created)



DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE Cadherin-related tumor suppressor homolog precursor (Fat protein  
 DE homolog).  
 GN FAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymphocytes;  
 RX MEDLINE=96163873; PubMed=8586420;  
 RA Dunne J., Hanby A.M., Poulsson R., Jones T.A., Sheer D., Chin W.G.,  
 RA Da S.M., Zhao Q., Beverley P.C.L., Owen M.J.;  
 RT "Molecular cloning and tissue expression of FAT, the human homologue  
 RT of the Drosophila fat gene that is located on chromosome 4q34-q35 and  
 RT encodes a putative adhesion molecule.";  
 RL Genomics 30:207-223(1995).  
 CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY EPITHELIAL AND SOME  
 CC -1- ENDOTHELIAL AND SMOOTH MUSCLE CELLS.  
 CC -1- SIMILARITY: Contains 34 cadherin domains.  
 CC -1- SIMILARITY: Contains 5 EGF-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin G-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X87241; CA60685.1; -.  
 DR HSSP; P01132; IEGF.  
 DR Gene; HGNC:3595; FAT.  
 DR MIM; 600976; -.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0008181; F:tumor suppressor; TAS.  
 DR GO; GO:0007155; P:cell adhesion; TAS.  
 DR GO; GO:0007267; P:cell-cell signaling; TAS.  
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF 2.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR001791; Laminin-G.  
 DR Pfam; PF00028; cadherin\_33.  
 DR Pfam; PF00008; EGF\_5.  
 DR Pfam; PF00054; laminin\_G\_1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 31.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00282; LamG\_1.  
 DR PROSITE; PS00010; ASX HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 16.  
 DR PROSITE; PS00268; CADHERIN\_2; 33.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS00025; LAM G DOMAIN; 1.  
 KW Cell adhesion; Signal; Glycoprotein; Transmembrane; Calcium-binding;  
 KW Repeat; EGF-like domain.  
 FT SIGNAL.  
 FT CHAIN.  
 FT 1 21 POTENTIAL.  
 FT 22 4590 CADHERIN-RELATED TUMOR SUPPRESSOR  
 FT 22 4183 HOMOLOG.  
 FT TRANSMEM 4184 4204 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 4205 4590 POTENTIAL.  
 FT DOMAIN 22 149 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN CADHERIN 1.

FT 150 256 CADHERIN 2.  
 FT 257 361 CADHERIN 3.  
 FT 362 463 CADHERIN 4.  
 FT 464 569 CADHERIN 5.  
 FT 570 716 CADHERIN 6.  
 FT 717 822 CADHERIN 7.  
 FT 823 927 CADHERIN 8.  
 FT 928 1034 CADHERIN 9.  
 FT 1035 1138 CADHERIN 10.  
 FT 1139 1245 CADHERIN 11.  
 FT 1246 1345 CADHERIN 12.  
 FT 1346 1456 CADHERIN 13.  
 FT 1457 1562 CADHERIN 14.  
 FT 1563 1670 CADHERIN 15.  
 FT 1671 1769 CADHERIN 16.  
 FT 1770 1882 CADHERIN 17.  
 FT 1883 1982 CADHERIN 18.  
 FT 1983 2084 CADHERIN 19.  
 FT 2085 2185 CADHERIN 20.  
 FT 2186 2286 CADHERIN 21.  
 FT 2287 2393 CADHERIN 22.  
 FT 2394 2495 CADHERIN 23.  
 FT 2496 2599 CADHERIN 24.  
 FT 2600 2705 CADHERIN 25.  
 FT 2706 2811 CADHERIN 26.  
 FT 2812 2920 CADHERIN 27.  
 FT 2921 3024 CADHERIN 28.  
 FT 3025 3127 CADHERIN 29.  
 FT 3128 3232 CADHERIN 30.  
 FT 3233 3337 CADHERIN 31.  
 FT 3338 3442 CADHERIN 32.  
 FT 3443 3546 CADHERIN 33.  
 FT 3547 3649 CADHERIN 34.  
 FT 3790 3827 EGF-LIKE 1.  
 FT 3831 4011 LAMININ G-LIKE.  
 FT 4013 4051 EGF-LIKE 2.  
 FT 4052 4089 EGF-LIKE 3.  
 FT 4090 4126 EGF-LIKE 4.  
 FT 4127 4163 EGF-LIKE 5.  
 FT 4164 4333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4334 4400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4401 4468 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4469 4536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4537 4604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4605 4672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4673 4740 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4741 4808 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4809 4876 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4877 4944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 4945 5012 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5013 5080 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5081 5148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5149 5216 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5217 5284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5285 5352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5353 5420 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5421 5488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5489 5556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5557 5624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5625 5692 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5693 5760 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5761 5828 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5829 5896 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5897 5964 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 5965 6032 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6033 6100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6101 6168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6169 6236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6237 6304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6305 6372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6373 6440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6441 6508 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6509 6576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6577 6644 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6645 6712 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6713 6780 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6781 6848 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6849 6916 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6917 6984 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 6985 7052 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7053 7120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7121 7188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7189 7256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7257 7324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7325 7392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7393 7460 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7461 7528 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7529 7596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7597 7664 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7665 7732 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7733 7800 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7801 7868 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7869 7936 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 7937 8004 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8005 8072 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8073 8140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8141 8208 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8209 8276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8277 8344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8345 8412 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8413 8480 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8481 8548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8549 8616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8617 8684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8685 8752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8753 8820 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8821 8888 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8889 8956 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 8957 9024 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9025 9092 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9093 9160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9161 9228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9229 9296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9297 9364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9365 9432 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9433 9500 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9501 9568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9569 9636 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9637 9704 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9705 9772 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9773 9840 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9841 9908 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9909 9976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 9977 10044 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10045 10112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10113 10180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10181 10248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10249 10316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10317 10384 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10385 10452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10453 10520 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10521 10588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10589 10656 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10657 10724 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10725 10792 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10793 10860 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10861 10928 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 10929 11000 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11001 11068 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11069 11136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11137 11204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11205 11272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11273 11340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11341 11408 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11409 11476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11477 11544 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11545 11612 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11613 11680 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11681 11748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11749 11816 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11817 11884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11885 11952 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 11953 12020 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12021 12088 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12089 12156 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12157 12224 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12225 12292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12293 12360 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12361 12428 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12429 12496 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12497 12564 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12565 12632 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12633 12700 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12701 12768 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12769 12836 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12837 12904 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12905 12972 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 12973 13040 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13041 13108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13109 13176 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13177 13244 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13245 13312 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13313 13380 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13381 13448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13449 13516 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13517 13584 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13585 13652 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13653 13720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13721 13788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13789 13856 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13857 13924 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 13925 14000 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14001 14068 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14069 14136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14137 14204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14205 14272 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14273 14340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14341 14408 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14409 14476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14477 14544 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14545 14612 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14613 14680 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14681 14748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14749 14816 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14817 14884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14885 14952 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 14953 15020 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15021 15088 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15089 15156 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15157 15224 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15225 15292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15293 15360 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15361 15428 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15429 15496 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15497 15564 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15565 15632 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15633 15700 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15701 15768 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15769 15836 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15837 15904 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15905 15972 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 15973 16040 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16041 16108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16109 16176 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16177 16244 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16245 16312 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16313 16380 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16381 16448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16449 16516 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16517 16584 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16585 16652 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16653 16720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16721 16788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16789 16856 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16857 16924 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16925 16992 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 16993 17060 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17061 17128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17129 17196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17197 17264 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17265 17332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17333 17400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17401 17468 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17469 17536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17537 17604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17605 17672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17673 17740 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17741 17808 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17809 17876 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17877 17944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 17945 18012 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18013 18080 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18081 18148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18149 18216 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18217 18284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18285 18352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18353 18420 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18421 18488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18489 18556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18557 18624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18625 18692 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18693 18760 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18761 18828 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18829 18896 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18897 18964 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 18965 19032 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19033 19100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19101 19168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19169 19236 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19237 19304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19305 19372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19373 19440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19441 19508 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19509 19576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19577 19644 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19645 19712 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19713 19780 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19781 19848 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19849 19916 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19917 19984 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 19985 20052 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20053 20120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20121 20188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20189 20256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20257 20324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20325 20392 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20393 20460 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20461 20528 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20529 20596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20597 20664 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20665 20732 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20733 20800 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20801 20868 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20869 20936 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 20937 21004 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21005 21072 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21073 21140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21141 21208 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21209 21276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21277 21344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21345 21412 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21413 21480 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21481 21548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21549 21616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21617 21684 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21685 21752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21753 21820 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21821 21888 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21889 21956 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 21957 22024 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22025 22092 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22093 22160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22161 22228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22229 22296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22297 22364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22365 22432 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22433 22500 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22501 22568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22569 22636 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22637 22704 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22705 22772 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22773 22840 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22841 22908 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22909 22976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 22977 23044 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23045 23112 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23113 23180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23181 23248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23249 23316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23317 23384 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23385 23452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23453 23520 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23521 23588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 23589 23656 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT

Db 2497 QNEVEVLAENADPLHVLWEVKTDDSGIYGVTHI--VNDPANDRYINE--RGQIF 2552  
 QY 243 TAEELRETKSTYVF--RYTAQDHGMPRBSALATLFLVTDVTDHDPVFEQCKESLRE 300  
 Db 2553 TLEKLDREFPAEAVISRLMAKQAG--GKVAFCVAVIITDDDNAPQFATYEVANIGS 2610  
 QY 301 NLEVGVEVLTARATDGDAPPNANILYRLLEGSGSPSEVEIDRSGVITRGPVDRREV 360  
 Db 2611 SAAKGSIVVC--SASDDEGSDNADITY--AIBADESEVENLEINKLGSVITTKESLIGLEW 2668  
 QY 361 ESYQVLTASDQGRDQPRSTTAAPLVSVEDDDNNAPOPEKRYVQVQVEDYTPGAPVLR 420  
 Db 2669 EFTTFVRADNNSP--SKESVVLVYVKIIPPEMQLPKFSEPPYTTVSDVAVGIEIDL 2726  
 QY 421 VTASDRDKSNAAVHYSMSGNA-----RGQFLDAQGTALDVVSPLDYETK--EYTLR 473  
 Db 2727 IRBE-----HSGVVLXSVKGNTPESNRDSFVIDQSGRLKLEKLDHETTKYOPSLI 2781  
 QY 474 VRAQDGRPLSVNSGLVTVQVLDINDNAPIVSTPPQATVLESVPLGVILHVOAIDAD 533  
 Db 2782 ARCTQDHEWVASVD--VSIQVADANDNSPVFESSPEAFIVENLPGSGRVIQIRASDAD 2839  
 QY 534 AGDNARLEAYL---AGVGHDPPTINNGCMISVALEDRERDYPYSPGVARADHGTTP-A 589  
 Db 2840 SGNGVWMTSLDOSQSEVIESFAINMETGMITTLKELDHEKRDYQIKVVASDHCKIQ 2899  
 QY 590 LITASASVTVLVDVNNNPFTQPEYTVRLNEDAAVGTSVVTSAYDRADHAY---ITYQ 646  
 Db 2900 LSTLAVDVTVVDVNSPREFTAIRKGYSEDDPGGVYAIISTDDASEIRNQYITF 2959  
 QY 647 ITSGTRNRPSTISQSGGLVSLALPLDYKLERQVYLAVATSDGTRODTAQIVVNTDAN 706  
 Db 2960 ITGGDELCPFAVFTIONEMKVYVKEKLEDEKRDNYLITTAQGTSSKAIYEVKULADAN 3019  
 QY 707 THRPVQSSHVTYVNWEDRPAGTTVVLISATDEDTENARITY--FHEDSIPQRIDADNG 765  
 Db 3020 DNSPVEKTLSTPTIEPDVLPGLIMQISATDADISNABITYTLGSGAEKRPALPDG 3079  
 QY 766 AVTTQALEYEDQVSYTLAITARDNG--IPQSDTYLELIVVDVNDNAPQFLRDSYQGS 823  
 Db 3080 ELKTSPLRREBAVHLVLRATDGGRRPQASIVTLE---DVNDNAPERSADYAT 3135  
 QY 824 VVEDVPFVSVAQISATDRDSCINGRVTTPQGGDGDGDFVIESVSGIVTRLRIRDEN 883  
 Db 3136 VVENETPGLTLRVQATDADAGIKRLVSLI--DBADQGFISINELSGIIQLEKPLDREL 3193  
 QY 884 VAGYVRAVAVDKMPARTPM--EYTVTVLDVNDNPVFEQDBFDVFEVENSPIGAVAR 942  
 Db 3194 QAVYTLSLKAVDQGLFRRLTAGTIVASVLDINDNPVEYREYGAIVSEBILVGTGLQ 3253  
 QY 943 VTATDEGTAQIMTQIVEGNIPEVFOIDIFSGELTAVLDLDEDRPEVYLVIOAT--- 999  
 Db 3254 VYAAASDIENAMETYSIISGNHKGFSIDSKTGAFFIENDYBESHYYLVLEATIDG 3313  
 QY 1000 ----- 999  
 Db 3314 TPEISDAVAVNVATDINDNTPVFSQDTYTVTISEDVALEQSVITWADADGSPNSHIH 3373  
 QY 1000 -----SAPVSRATYHVRLL 1014  
 Db 3374 YSIIQNGQSSFTIDPVRGKVTKLLDRETISGYTLVQASDNGSPPVNNTTVINDIS 3433  
 QY 1015 DRNDNPVL--GNFELFNN----- 1032  
 Db 3434 DVANDNAPVSRGNYSVIIDENKPVGSVQLVTVBDSHNGPPFFITVGNDEKAFEV 3493  
 QY 1033 -----Y 1033  
 Db 3494 NPOGVLLTSSAIRKKEKHVLLQVKAADNGKPOLSSLYIDIRVIEESIYPALILEIF 3553  
 QY 1034 VTNRSASPFGAIGRAVADPDIDSLTYSFERGNELSLVILNASTGELKLSALDNNRP 1093  
 Db 3554 ITSSGEYSGGVIGKIHATQDVYDVTITYSIDPQMD-NLPSVSTGGKLIANKKLDIG-- 3610

QY 1094 LEAINSVLSDGSHVSTQACALRVTLITDEMLTHSITLLEDMSPERPLSLGLFIQAV 1153  
 Db 3611 -OYLAVSTVDBKFTTADITVHIROVQEMNLHTIALIFANULTEEFVGDYWRNFRL 3669  
 QY 1154 AATLATPPDHVVVFNVDRTDAPGGHILNVSLVSGPPGPGGPPPLPSEDLQERLYLNR 1213  
 Db 3670 RNILGVRARDIQIVSLQ--SSEPHH-LDVLLFVEKP-----GSAQISTKQLHKINSV 3721  
 QY 1214 SULTIASAQRVLPFDNITCLREPCENYKCVSLAPDSAPFIASSVL--FRPLHPVG 1271  
 Db 3722 TDIEHIGVRLINVPQKICAGIDCP--KRCDEKAVDSVMSHTSTARLSFVTPRHRRA 3780  
 QY 1272 LRC-----RCPPGFTGDCETEVDLCYSPGPHRCRSP--EGGYTCLCRDGYGENHE 1324  
 Db 3781 V-CLCKEGRCPVNHG-----CEDDPCFGESECVSDPHEKHCTVCPSG----- 3823  
 QY 1325 VSARSGRCTPGVCKXNGGTCVNLVGFKDCDPSGDPBEKPYCQVTRSPRAPHFITER--G 1382  
 Db 3824 ---RFGQC-PG-----SSGMTLIGNSVYKXRLTE 3848  
 QY 1383 LEQRHFTLASFATKENDGLLYNGRFRNEKDFVALSVIQVQLTBSAGSTTVSBF 1442  
 Db 3849 NENKLEMLTWMLRTYSTHAAVMY---ARGTVSILHHRQLQYKFCQSG----- 3897  
 QY 1443 VEGVS-----DQOMHTVOLKTYNKPRLGQYGLPQPSBQKAVVTVVDCDGVALLF 1495  
 Db 3898 -PGIYSVSGIQVNDQMANVALL----- 3919  
 QY 1496 GSVLGNYS-----CAAQGTQSGSKSLDTGPLLGGVDPDPSFVYMRQ----- 1541  
 Db 3920 --VNGNVARLVLDQVHTASGAPGLKTLINDNYVFFG-----HROQSTRHR 3967  
 QY 1542 -----FVCGM-----RULQVDSKHIDMADFLANNGYPGC--PAKANVCBNTGN 1585  
 Db 3968 SPQVNGFRGCWDSIYLNGQLPLNSKRSYAHIEESVDVSGCFLTIED--CASNPCCN 4026  
 QY 1586 GGTCC--VNQMDAFSCBPLFGFGKSCAQEMANPQHPLGSSLVAMHGLSPIQPYLSLAF 1644  
 Db 4027 GGVCNPSBAGGYCKCSALYIGTIC----- 4051  
 QY 1645 RTRQADGVLLAATRGRSTITLQLRBHVMLSVBSTGLQASSLRLEPGRANDGDMHNAOL 1704  
 Db 4052 -----EISVNPCSSN----- 4061  
 QY 1705 ALGASGGRHAILSPDYQGRABGNLGRHLHLSNITVGGIPEPAGVARGFGCIG 1764  
 Db 4062 -----PCLYG-----GTCVVDNGAFVCCQCKGLYTG 4086  
 QY 1765 VRVSDTPRGVNSLDPSSHGSIINVEQCSGLPDCDSNPCCANSYCSNDMSYSQCDPGY 1824  
 Db 4087 -----QRCOLSPYCXDEPRCKNGTCDSDLDGAVCCQDSGFR 4122  
 QY 1825 GDNC---TNVCDLNPCHQSVCTRKPSAPHG-YTCECPNLYGPCE 1867  
 Db 4123 GERCOSDIDECSGNPLGALCEN---THGSYHNCSEHYRGRHCE 4165  
 RESULT 12  
 FAT2\_DROME STANDARD; PRT; 4705 AA.  
 AC Q9VW71; Q95S51;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Putative fat-like cadherin-related tumor suppressor homolog  
 DE precursor.  
 GN FAT2 OR CG7749.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 OK NCBI\_taxid=7227;

[1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanaides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 Brannon R.C., Rogers Y.-H.C., Blazey R.G., Chame C.R., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.W., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Bernstein B.P., Bhandari D., Bolashkov S.,  
 Botkova D., Botchan M.A., Bouck J., Brockstein P., Brotter P.,  
 Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glaeser K.,  
 Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,  
 Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Liako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Mekhalov G., Misha N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
 Palaztolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svraksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 [2]  
 RN REVISIONS.  
 RP STRAIN=Berkeley;  
 RC MEDLINE=22426069; PubMed=12537572;  
 RX Miara S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,  
 Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 Beutenhout B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 Lewis S.E.;  
 "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 systematic review.";  
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 [3]  
 RN SEQUENCE OF 3837-4705 FROM N.A.  
 RP STRAIN=Berkeley; TISSUE=Ovary;  
 RC MEDLINE=22426066; PubMed=12537569;  
 RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Chame M.,  
 George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 Rubin G.M., Celniker S.E.;  
 "A *Drosophila* full-length cDNA resource.";  
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RL -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- SIMILARITY: Contains 34 cadherin domains.  
 CC -1- SIMILARITY: Contains 5 EGF-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin G-like domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AEO03515; AAF9078.2; -  
 DR EMBL: AY060955; AAL28503.1; ALT INIT.  
 DR EMBL: AY118666; AAM50035.1; ALT\_INIT.  
 DR HSSP: P15116; INCI.  
 DR FlyBase: FBgn0036930; fat2.  
 DR GO: GO:0005887; C:integral to plasma membrane; ISS.  
 DR GO: GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; ISS.  
 DR GO: GO:0016339; P:calcium-dependent cell-cell adhesion; ISS.  
 DR InterPro: IPR000152; Asx hydroxyl.  
 DR InterPro: IPR002126; Cadherin.  
 DR InterPro: IPR001881; EGF Ca.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR006210; IEGF.  
 DR InterPro: IPR001791; Laminin G.  
 DR Pfam: PF00028; cadherin; 31.  
 DR Pfam: PF00008; EGF; 5.  
 DR Pfam: PF00054; laminin G; 1.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 34.  
 DR SMART: SM00282; Lang; 1.  
 DR PROSITE: PS00010; ASX HYDROXYL; 1.  
 DR PROSITE: PSS0268; CADHERIN\_2; 34.  
 DR PROSITE: PSS0022; EGF\_1; 5.  
 DR PROSITE: PSS0186; EGF\_2; 2.  
 DR PROSITE: PSS0187; EGF CA; 1.  
 DR PROSITE: PSS0025; LAM G DOMAIN; 1.  
 KW Hypothetical protein; Cell adhesion; Signal; Glycoprotein;  
 KW Transmembrane; Calcium; Calcium-binding; Repeat; EGF-like domain.  
 FT CHAIN 1 35  
 FT 36 4705  
 FT 1647 1668  
 FT 1669 4705  
 FT 60 180  
 FT 181 288  
 FT 285 397  
 FT 398 504  
 FT 505 610  
 FT 611 713  
 FT 710 874  
 FT 875 977  
 FT 978 1085  
 FT 1086 1195  
 FT 1191 1296  
 FT 1297 1402  
 FT 1405 1503  
 FT 1504 1609  
 FT 1610 1714  
 FT 1715 1812  
 FT 1813 1929  
 FT 1929 2465  
 FT 2466 2567  
 FT 2568 2670  
 FT 2671 2779  
 FT 2780 2876  
 FT 2877 2983  
 FT 2984 3088  
 FT 3084 3185  
 FT 3186 3289  
 FT 3290 3394  
 FT 3395 3499  
 DOMAIN 36 1647  
 TRANSSEM 1648 1668  
 DOMAIN 1669 4705  
 DOMAIN 60 180  
 DOMAIN 181 288  
 DOMAIN 285 397  
 DOMAIN 398 504  
 DOMAIN 505 610  
 DOMAIN 611 713  
 DOMAIN 710 874  
 DOMAIN 875 977  
 DOMAIN 978 1085  
 DOMAIN 1086 1195  
 DOMAIN 1191 1296  
 DOMAIN 1297 1402  
 DOMAIN 1405 1503  
 DOMAIN 1504 1609  
 DOMAIN 1610 1714  
 DOMAIN 1715 1812  
 DOMAIN 1813 1929  
 DOMAIN 1929 2465  
 DOMAIN 2466 2567  
 DOMAIN 2568 2670  
 DOMAIN 2671 2779  
 DOMAIN 2780 2876  
 DOMAIN 2877 2983  
 DOMAIN 2984 3088  
 DOMAIN 3084 3185  
 DOMAIN 3186 3289  
 DOMAIN 3290 3394  
 DOMAIN 3395 3499  
 PUTATIVE FAT-LIKE CADHERIN-RELATED TUMOR  
 SUPPRESSOR HOMOLOG.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 CADHERIN 1.  
 CADHERIN 2.  
 CADHERIN 3.  
 CADHERIN 4.  
 CADHERIN 5.  
 CADHERIN 6.  
 CADHERIN 7.  
 CADHERIN 8.  
 CADHERIN 9.  
 CADHERIN 10.  
 CADHERIN 11.  
 CADHERIN 12.  
 CADHERIN 13.  
 CADHERIN 14.  
 CADHERIN 15.  
 CADHERIN 16.  
 CADHERIN 17.  
 CADHERIN 18.  
 CADHERIN 19.  
 CADHERIN 20.  
 CADHERIN 21.  
 CADHERIN 22.  
 CADHERIN 23.  
 CADHERIN 24.  
 CADHERIN 25.  
 CADHERIN 26.  
 CADHERIN 27.  
 CADHERIN 28.  
 CADHERIN 29.  
 CADHERIN 30.  
 CADHERIN 31.  
 CADHERIN 32.

```

FT DOMAIN 3500 3604 CADHERIN 33.
FT DOMAIN 3605 3712 CADHERIN 34.
FT DOMAIN 3819 3879 EGF-LIKE 1.
FT DOMAIN 3881 3919 EGF-LIKE 2.
FT DOMAIN 3937 4121 LAMININ G-LIKE.
FT DOMAIN 4129 4166 EGF-LIKE 3.
FT DOMAIN 4168 4205 EGF-LIKE 4.
FT DOMAIN 4243 4279 EGF-LIKE 5.
FT DISULFID 3823 3835 POTENTIAL.
FT DISULFID 3830 3867 POTENTIAL.
FT DISULFID 3869 3878 POTENTIAL.
FT DISULFID 3885 3896 POTENTIAL.
FT DISULFID 3890 3907 POTENTIAL.
FT DISULFID 3909 3918 POTENTIAL.
FT DISULFID 4133 4144 POTENTIAL.
FT DISULFID 4138 4154 POTENTIAL.
FT DISULFID 4156 4165 POTENTIAL.
FT DISULFID 4172 4183 POTENTIAL.
FT DISULFID 4177 4193 POTENTIAL.
FT DISULFID 4195 4204 POTENTIAL.
FT DISULFID 4247 4258 POTENTIAL.
FT DISULFID 4252 4267 POTENTIAL.
FT DISULFID 4269 4278 POTENTIAL.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 923 923 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1198 1198 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1312 1312 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1473 1473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1511 1511 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 3962 3962 G -> E (IN REF. 3; AAL28503).
SQ SEQUENCE 4705 AA; 524564 MW; 6D387A489D2C3DR CRC64;

```

Query Match 9.5%; Score 1477.5; DB 1; Length 4705;  
 Best local similarity 22.8%; Pred. No. 1.9e-62;  
 Matches 462; Conservative 282; Mismatches 693; Indels 593; Gaps 53;

```

QY 178 APOPOPSIOATYPRNOPAGTPVASLRALIDPDGEAGRLTYMDALPDRSNQFSLDPV 237
DB 2564 SPFDOSTYEAQVPEMLHGHNIITVAKASDGDGTANYLT---EIVSEMKKIFLIDQT 2620
QY 238 TGAUTTAELDRRTKSTHVRVTAADHGMFRSALATLTLYTDTNDHDPVEHQGQKXS 297
DB 2621 TGVITKVTFRKQKDEYVVLKVSDDG--GKFGFASLKVTVDVNDVPEFLKEIKXV 2678
QY 298 LRENLEVGVEVLTVRAIDGDAPNPANILYRLLEGSGG--SPSEVFEIDPRSGVIRTRGPV 356
DB 2679 VSTTYEANGQITLVKAKDDIDVNGSVHFGIYOKSKMDKAVKDIETNEKTGDIVFPSKXE 2738
QY 357 REEVSEYQLTVKASDGR-----DPE--- 377
DB 2739 SYGVNSYQFVRSADSGEPQFHSSEVPSIETDANIPFKFSVYLKIIESTPRTYVL 2798
QY 378 -----PRSTTAVF--- 386
DB 2799 TTKHMITGNITFKSIAADQHFMIISGELLLOOTLDRQESHNLIIVAEISTVVFPA 2858
QY 387 -----LSVEDNDNAPQFSEKRYVVOVREDVTPGAPVLRTVADSRRKSNAAVHVSIMG 441
DB 2859 YADVLDIVDRNDNYPKFDWTFYSASVAENSEKVISLVKVSATDADDTGPGDIRRYLSE 2918
QY 442 --NARQGYLDAQTGLDVSPIDYETTKYTLRVAADGGRPLSNVSGLVTVQVLDIN 499
DB 2919 TENIQIFIDIDISGITLTSLDREVQSEYKFNVAADNGHR--HDARVPYTIKIVDYN 2977
QY 500 DNAPIVSTFPAQ--TVLESVPLGLVYLVAQIDADAGDNARLEFRLAGVGHDPFTING 558
DB 2978 DNAPVF--KLPISGLSVEENALPQTIVLINTLLIDPDI--EKQEMDFIVSGDKQAQFOIGK- 3034

```

```

QY 559 TGMISVAELDEBEVDVFSGYEARDHGTPLTASASVTVLVNDNNPTFQPEYTVR 618
DB 3035 SGEPLFAKLDEBQMLFNLSTIAD--GKTLANAVIEVDINDNPPYCKRPHHS 3091
QY 619 LNEDAVGTSVTVASANDDAHSVITTYQITSGTRNRFISITSGGGLVSLALPDYKLE 678
DB 3092 TRESISTGTLVEKVIDFDPOSKLRFYL--SGKADDFSIGES--GILKVASALDRETT 3148
QY 679 ROYVLAVTASDQ---TRODTAQIVNVTWANTHRPFGSSHVTYVNERPAGTIVLIS 735
DB 3149 PKYIVAHVODGDPTQBCFSEIITVNDINDNMPFSAQYRVSPEDAQIATLITKYH 3208
QY 736 ATDEDEGENARITYFM----- 751
DB 3209 AWDKDGVRKQIKTSLMGENDHYFKISKTGILRLKSLDRETIISLFLNLTVAEDCGVR 3268
QY 752 -----BDSIQ----- 757
DB 3269 LNSIATVAVNIDINDNPPFPMRQYSCKILENATGTEVCKYVATSIDGVNADHYFI 3328
QY 758 -----PRIDADGAVTTOAEIDYEDOVSTLATAIDNGIPQKSDTYLEILVNDV 809
DB 3329 MGENQKFKMDSTGDLVLAATLIDYEMSKFYELTQALDGGTPPLSNNAVYNISILDIN 3388
QY 810 DNAPQFLRDSYQSYVEDVPFTSVLQISATDRDGLNGRVYTRFGQDGDGFVEST 869
DB 3389 DMSPTFLQVLYANNEDIPVSGKILDVATITBEDSDVNGLVYNNIRGN--IGQSIDK 3447
QY 870 SGIVTLRLRENNVAQVYLAAYADKMPARTMEVTVTVLVNDNPPVEODEFDVF 929
DB 3448 NCTISGRPLDRETTISHTYLEIQACDQGPQCSNVPININILDTNDNAPIFSSNYSV 3507
QY 930 VENSPIGLAVARVATPDDEGTN--AQIMQVBEINPEVPODIFSGELTALVDLYD 988
DB 3508 IQENRLIGVLTFRKISDDETPNTPTPTDRSNBEGGLFRL--QDGLRTASRNNHL 3566
QY 989 REEYVL---VIGTASPLVRSATVAVRLDRNDNPVLGNFELFNNTYTNSSSPGGA 1045
DB 3567 QDEPVIQVRVPNGRPPLYSADMVVKIIEBQYPIVPLEVTINSF---EDPESGAF 3622
QY 1046 IGRVAPHDIDISLTYSFERGNE--LSVLILANS--TGBIKLSPALDNNRPLEAINSV 1100
DB 3623 ICKVHASQDKYDEINFSLVSGPDMDYOSKLFNINNTYTKIYASINDIGL--YKLVN 3679
QY 1101 LVSQGHVSTACALRVITIDEMLTSLTLEMSPERFISPLIGLQAVATIANP 1160
DB 3680 SVSDGKFTVFSIVKINVELVINDMLKESVIRPRIISAEFLSHRKTFRSIRNIMCR 3739
QY 1161 PDHVVVFNVDRTDAPGHI-----LNVSLSVGPPGPGGPPFLPSEDLOR 1208
DB 3740 QKDVLITLQSDYQKASQAHAVGNRPARSIDSLNVVPAVRKQIILPDSBFSTSEIRQT 3799
QY 1209 LYIARSL--TALSAQVLP---FDNITCLAREPCENYMCVS--VLRPSSAPFLAS 1258
DB 3800 LIDKNGEIEENENLVVEBVLSTCOXNKNDCHGCKOILQILKNVYTTFTFDVISAAP 3859
QY 1259 SSVLFRPHVPGGLRCRPGTGDYCEYTVDLCVSRGCPGRGRSREG--GYCLORD 1316
DB 3860 SYI-----PVN--TCVCRPGDGRHCKEYVACSTDPCSFRICMPGSGALGYCVCCK 3911
QY 1317 GYTGHECVSASRGTCPGVCNKGCTCVNLVVGFCPCPSGDPFEPYCOVTRTSFPNHS 1376
DB 3912 GFSGTyce--RKSXCSNBSOMG-----LPTAVSFGGS 3944
QY 1377 FTFRGLAORFHTL-----ALSPATKROGLLYNGRFNEKHDPAVLETVQEOVLT 1430
DB 3945 YAHYKINKAKAFTLENGFSYSIQIRTVQGTGLLY---ASQKDYNNLEIINGAVQRF 4001
QY 1431 --SAGESTTVSPFVPGVSDQOMHTVOLKYNNKLLQOTGLPQSPSOKAVAVTVGDCD 1488
DB 4002 DLSGSEGVISYST---NISDGMHQISLER----- 4029

```

QY 1489 TGVALRFSVLTGANTSCAAQGTGGSKSLDTLTPILLGCVDPDLE---SPFVNRQFVG 1544  
 Db 4030 ---SINSKAKVWMDNGVHSGAPGVNGLINTQSDNIFVGAERVPHPSITIGEDIQRFIG 4086  
 QY 1545 GGRNLOVDSRHDMAADFIANNGTVPACPAKXV---CDSENTCHNGTGC 1589  
 Db 4087 CMAIKIKAKESLPL--YISGSSTIALAKRFTVNEKCPDPSNVLVLGICSSQPCANSGTIC 4144  
 QY 1590 VN-QMDAFSCCEPLGFGKSCQAQEWANPOHFLYSSSLVAMHGLSLPISQPMWLSIMFRTQ 1648  
 Db 4145 KELDTVECAACQPRYSGKHCEIDL----- 4159  
 QY 1649 ADGVLLOAITGRSRTITQLRGRHVMSVEGTGLQASLRLEPRGRANDGMHQAALGA 1708  
 Db 4170 ----- 4169  
 QY 1709 SGGFHAITSFYDGGQRAEGLNLPRLHGLHLSNITVGI PGPAGVGANGFGCQGVVVS 1768  
 Db 4170 ----- 4169  
 QY 1769 DTPEGVNSLDPSHGESINVEQCSLPDPCDSNPCPANSYCS-NMWSYSCSCDPRYVDN 1827  
 Db 4170 -----DPCSSGSPCLFGRCQDYHGNPNYSCTCPHLSGKR 4203  
 QY 1828 CT--NVCIDINPCHEQSVCTRKPASAPHGTYTCECPNYLGPYCTETRID---QPCRGMWGH 1861  
 Db 4204 CEYGFCTCPNCPKNGKGLCEBDGISH---CMC-RGYTGPTEBIDVDECENQPCNGG--- 4255  
 QY 1862 PTCGFCNCDVSKGFPDPCNKTSGECHKENHYRPPGSPYTCCLCOCYPTGS 1931  
 Db 4256 ATC-----INERBSFRC-IC8YTLTGA 4276

## RESULT 13

CADN DROME STANDARD; PRT: 3097 AA.  
 AC O15943: Q9VJ37;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neural-cadherin precursor (Cadherin-N protein) (DN-cadherin).  
 GN CADN OR CG7100.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM D).  
 RC TISSUE=Embryo, and Head;  
 RX MEDLINE=97388431; PubMed=9247265;  
 RA Iwai Y., Utsu T., Hirano S., Steward R., Takeichi M., Uemura T.;  
 RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion  
 RT receptor, in the Drosophila embryonic CNS.";  
 RL Neuron 19:77-89(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baxu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan D.A., Bouck J., Brokstein P., Brotter P.,  
 RA Buttle K.C., Busam M.R., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fodor C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glöckler A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin M., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemblous J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Melnikov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard V., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Snyder R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrstkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Berkely;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Belletcourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Ruben S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP INTERACTION WITH ARM.  
 RX MEDLINE=98298928; PubMed=9635189;  
 RA Loureiro J., Pfeiffer M.;  
 RT "Roles of Armadillo, a Drosophila catenin, during central nervous  
 RT system development.";  
 RL Curr. Biol. 8:622-632(1998).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL  
 CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL  
 CC INFORMATION.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=8;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=D;  
 CC IsoId=O15943-1; Sequence=Displayed;  
 CC Name=A;  
 CC IsoId=O15943-2; Sequence=VSP\_000667, VSP\_000668;  
 CC Name=B;  
 CC IsoId=O15943-3; Sequence=VSP\_000668;  
 CC Name=C;  
 CC IsoId=O15943-4; Sequence=VSP\_000667, VSP\_000668, VSP\_000669;  
 CC Name=E;  
 CC IsoId=O15943-5; Sequence=VSP\_000667;  
 CC Name=F;  
 CC IsoId=O15943-6; Sequence=VSP\_000669;  
 CC Name=G;  
 CC IsoId=O15943-7; Sequence=VSP\_000667, VSP\_000669;  
 CC Name=H;  
 CC IsoId=O15943-8; Sequence=VSP\_000668, VSP\_000669;  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN

CC THE MESODERM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE  
CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE  
CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN  
CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN  
CC THE CNS NEUROPILE, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT  
CC MUSCLES.  
CC -1- SIMILARITY: Contains 16 cadherin domains.  
CC -1- SIMILARITY: Contains 3 EGF-like domains.  
CC -1- SIMILARITY: Contains 2 laminin G-like domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB002397; BAA23151.1; -  
CC EMBL; AE003656; AAF53635.1; -  
CC EMBL; AE003656; AAN10992.1; -  
CC EMBL; AE003656; AAN10993.1; -  
CC EMBL; AE003656; AAN10994.1; -  
CC EMBL; AE003656; AAN10995.1; -  
CC EMBL; AE003656; AAN10996.1; -  
CC EMBL; AE003656; AAN10997.1; -  
CC EMBL; AE003656; AAN10998.1; -  
CC PIR; T00021; T00021.  
CC HSSP; P00740; 1EDM.  
CC FlyBase; FBgn015609; Cadr.  
CC GO; GO:0005911; C:intracellular junction; IDA.  
CC GO; GO:0008014; F:calcium-dependent cell adhesion molecule ac. .; IPI.  
CC GO; GO:0007412; P:axon target recognition; IMP.  
CC GO; GO:0007156; P:homophilic cell adhesion; IDA.  
CC InterPro; IPR00233; Cadherin.  
CC InterPro; IPR000742; EGF\_2.  
CC InterPro; IPR001881; EGF\_Ca.  
CC InterPro; IPR006209; EGF-like.  
CC InterPro; IPR001791; Laminin\_G.  
CC Pfam; PF00028; cadherin; 14.  
CC Pfam; PF01049; Cadherin\_C\_term; 1.  
CC Pfam; PF00008; EGF; 3.  
CC Pfam; PF00054; laminin\_G; 2.  
CC PRINTS; PR00205; CADHERIN.  
CC SMART; SM00112; CA; 16.  
DR SMART; SM00179; EGF CA; 1.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00232; CADHERIN\_1; 9.  
DR PROSITE; PS0268; CADHERIN\_2; 16.  
DR PROSITE; PS00022; EGF\_1; 3.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS00025; LAM G DOMAIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
KW Signal; EGF-like domain; Alternative splicing.  
FT SIGNAL 1 36  
FT PROPEP 37 ?  
FT CHAIN ? 3097  
FT DOMAIN ? 1454  
FT TRANSMEM 1455 1475  
FT DOMAIN 1476 3097  
FT DOMAIN 181 305  
FT DOMAIN 430 543  
FT DOMAIN 554 651  
FT DOMAIN 660 756  
FT DOMAIN 766 858  
FT DOMAIN 867 968  
FT DOMAIN 978 1078  
FT DOMAIN 1087 1183  
FT DOMAIN 1193 1299  
FT DOMAIN 1307 1414  
FT DOMAIN 1423 1514  
FT DOMAIN 1523 1630

FT DOMAIN 1639 1742 CADHERIN 13.  
FT DOMAIN 1749 1861 CADHERIN 14.  
FT DOMAIN 1870 1966 CADHERIN 15.  
FT DOMAIN 1974 2085 CADHERIN 16.  
FT DOMAIN 2146 2277 EGF-LIKE 1.  
FT DOMAIN 2379 2585 LAMININ G-LIKE 1.  
FT DOMAIN 2592 2627 EGF-LIKE 2.  
FT DOMAIN 2631 2822 LAMININ G-LIKE 2.  
FT DOMAIN 2869 2902 EGF-LIKE 3.  
FT DISULFID 2346 2357 POTENTIAL.  
FT DISULFID 2351 2366 POTENTIAL.  
FT DISULFID 2368 2377 POTENTIAL.  
FT DISULFID 2592 2607 POTENTIAL.  
FT DISULFID 2601 2616 POTENTIAL.  
FT DISULFID 2618 2627 POTENTIAL.  
FT DISULFID 2869 2880 POTENTIAL.

Query Match 9.1%; Score 1419.5; DB 1; Length 3097;  
Best Local Similarity 24.7%; Pred. No. 6.3e-60;  
Matches 512; Conservative 288; Mismatches 753; Indels 521; Gaps 72;

QY 179 PQQPPS-YQATVPEQNPAGTPVSLRALIDPDEGAEGLRYTMALPDSRNOQFSLDPV 237  
DB 966 PVFQDCSTTPPKVEBGA PNGSPVIAKVATDEKGVNQVKYSIVQOPQKGTK-FTYDER 1024  
QY 238 TGAVTTAEELDRETKSTHFRVLT--AODHGMFRSALATLITVTDNDHDVPVEQOEYK 295  
DB 1025 TGEVSTNKKVFPDRBEDDKFVSVTYKATDQDPSLEGVCSFTVELTVDNDNPFLPDRKXY 1084  
QY 296 ESLHENEVGYEVLTVPATGDAFPNANILYRLIEGSGSBSVEFEIDPSGVIRTNGPV 355  
DB 1085 ENVIQDASIGTNILIRVASDEDDANNCAIVSLTAPFPNDLEFEIQAESGWIIVLKPL 1144  
QY 356 DREVESEYQLTVEASDQGRDGRPSTTAAPFLSYEDNDNAPQSEKRY-VQVREBYTP 414  
DB 1145 DR---ETTKLEANAQDGYR--PLSKTVEVQIDVDVDAANPEVDHTVYGEFTYKENVV 1199  
QY 415 GAPYLRYTASDRDKGSNAVVHYSIMSG---NARQGFYL-----DAQTGA-LDVVSPLD 463  
DB 1200 GGRVYSIKASGIEG-NPTVFRMLPGSTAGTNKFEHFFYLQQRDNDGTADIVNHPD 1258  
QY 464 YETTKETTLAVRADGGRPPLSNVSGLYTVQVTDINDNAPLPVSTPPQATYLSVPYGL 523  
DB 1259 YESIKENYLTIRVENNNAQOLAS-EATVYIMLEBVNDIEPLFTRE-QEYVLEBEPITGYK 1316  
QY 524 VLVQAIDADAG-DNARLEYRLA---GVGHDPFTINNGWISVAABEDREVDPSYF 578  
DB 1317 VTQVNAIDKDGTPPNQVYTYIYDSPRNEKEF-FEINLOSGLIFTTVFDRKKAYAL 1375  
QY 579 GVEARDHGTALFASAS-----VSATVLDVNDNPPFTQPEYTVRLNEDAAVGTSVV 630  
DB 1376 EVEARD-GAPSARNSNGSPSVTFIRIGIADKNDNPPYDKSYEARVDENEDIGHTVL 1434  
QY 631 TVSAVVDASVITYQTSTNRRRSITSQSGGGLSLALPLDYKLERQYLAVTASDG 690  
DB 1435 IVTAKDDESSRIYELTSGNIGAPAVKNMTGA--IYVAGALDYETRRRYELRLASDN 1492  
QY 691 TROPDIAIVNVVDANTHRPVPFOSHYTVANVE-----DRP----- 726  
DB 1493 LKENYTVIIVHKDVNDNPVFERPTRTQITTEDDNLPRVLAQVTATDGDNDPQNIIV 1552  
QY 727 ----- 726  
DB 1553 YFLTGQIDDPDNPANSKFDINRTTGIFVLKPLDRDQPNGRPQWRFTVFAQDEGEGVLNG 1612  
QY 727 -----AGTYVLISADDE-----GEMARITYEM 751  
DB 1613 YADVQVNLKIDINDNAPIFQGVYFGVNTENGTAQVMTAVDYDDPNBSNARLVYSI 1672  
QY 752 EDSI-----POFRIDADGAVTQA----- 771  
DB 1673 EKQVIEEFTOSPIFEIIPDTGVIAKTAVCCIDRERTPDYSIQVAMQGGGLKGTGTASIRV 1732

```

QY 772 ----- 771
Db 1733 KDINDMPOFTKDEWTEVDETDGTALPEMPLITVYHDEDETNKFQYKVIDNSGYADK 1792
QY 772 ----- 771
Db 1793 FTWVNNNDGTSGLKIVQPLDYEDQLOSNNGFRFRIOVNDKGBENDNDKXVAVSWVAVFLR 1852
QY 807 DVNDNAPQPLRDSYQGSVYEDVPFTSVLQISATDRDSGLNGRVYTFQGGDGDGDFIV 866
Db 1853 DINDNKPHEERANVSVFEDTKVTGLKPKATTPDQGSKSVYSIDRSSDRQRAI 1912
QY 867 ESTSGIVTLRLRDRNVAQVYLAAYVDKMPARTMEVTVTLVDVNDPVPVEQDEF 926
Db 1913 -NONGSVTIOBSLDEVVPRHQVKILAIIDGSPPKTATATLTVIYQINDNAPKELKDYR 1971
QY 927 DVFVENSPIGLAVARVATDPDEBCTNAQIMQIYEGNIPEVPOLDIISGE----- 977
Db 1972 PVLPEHVP--RKVEIILATDDDRSKS-----NGPPQFRLDPADDIIRASPFVE 2021
QY 978 ----- 1792
Db 2022 QDQKANDGMAVIVSLRSPDEQKQYMIPIVTDHSGSPAMTGTSTLTVIIGDVNDKM 2081
QY 1022 VLGNFELFNNTVTRSSSPFGAIGRVPAHD--DISDLTY-----SFERGELS 1071
Db 2082 QGSKDIFVYNY--QGQSPPTPIGRVYVVDLDMDLDPDKKFYEAMEHPRFKLDEDSG 2137
QY 1072 LVILNASTGE-----LTLSTRALDNNRLEAIMSVLVDGYH--SYTACALRVITTEM 1124
Db 2138 MVTMAGRGREGRYHLRFYVDRKQTQTDIPANVTYVREIPEAVNNGSVRLSISDED 2197
QY 1125 LTHSTTLLEDMSPERFLSPILGLEIQAVAATLAPPHVVFVNYQRTDAPGHIILNS 1184
Db 2198 FIVWNNYTGQMSNRK-----MDRFRDLADLINTERENVOIFSQQLRKRP--PLTVR 2250
QY 1185 LSV--GQP--PQGGGPPPLPSEDIQERLYANRSLTLAISQAVLPFPDNDICLRPCENY 1240
Db 2251 FSAHSGPYKPRVRLNGIYVLMHREIEKVGIN--ITWVGIDECI--YENQMC--BGSCTNS 2305
QY 1241 MRCVSLAFDSAPFIASSSVLFRITHVVGRLCRCPGFGDVCETEVDLCYSRCPBH 1300
Db 2306 LFI-----SPLFVWVNN-----KTALVGVAVDTIAD----- 2332
QY 1301 GRCSRREGYTCLCRDGYTGEHCEVSASGRCTPGVCKNGGTQVNLVGFQKDCPSGDF 1360
Db 2333 --C-----TCGANFTKPBSC-----RTTP--CHNGRCVDTFRPG--HCSFPGV--Y 2372
QY 1361 EKPVCQVTRSPRAHSFTTFRGLRORFHTLALSPATERDGLLYNGRFNEKH----- 1414
Db 2373 TGPRCQOQTRSPFRNGMAMYPLLEWCDSHLSLEFITTCKPDGLIYNGPIVPERDETLI 2432
QY 1415 -DFVALLEIOBOVOLTFASGSETTVSPFVPGVSDGOMHTOLKYKPKLLGQGLRPG 1473
Db 2433 SDFILBLEERGPRLIDPGSGTLELYKTKTLDGEMHRIDLEWDTESIR----- 2484
QY 1474 PSBOKVAVVYDGC-----DTGVALRFGSVLGNVSCAOGTQGSKKSLDLPULL 1525
Db 2485 -----MVVDFCASARIEAEMEDGTPEPFDN-----SQQARQGPPEVETLVNVAFLQV 2532
QY 1526 GGVPLPSSFPVYM-----RQFVGMRLQVDSRHIDMADIANNGVPGCPACKN 1576
Db 2533 GGL--YRQFQDQSLFWMHYMPAKPDCIRNLVNSKLYDLAHGLSRNSVAGCPQEB 2590
QY 1577 VC-----DBNTCHNGGTQVNMOMDAFCECPGLFGKSCAQMAMNPOHFGSSIVAHGSL 1632
Db 2591 VCAQETTRACMEHNCVGSLSHACHCRPGWTGAC--NIPITPTTAAQSVYKALSF 2649
QY 1633 PISQFWYLSLWFRTRQADGVLLQALTR--GRSTITLQLRGGV--MLSVETGLQASSLRL 1689
Db 2650 FDRFSTQVQLRFRTRBEYGEFLFRVSDQNHREYGLIIEKHGLHFRYNNLSRTEKDLML 2709
QY 1690 BPRANDDMHHAQLALASGGPRAIILSPYQQRABGN-----LGRRLHGLHL 1739

```

```

Db 2710 NAIIVNDGOMHVVVNNRYGSA-----ATLELDGEGRRYNETFEFVGHQMLLVNDQKQGVY- 2764
QY 1740 SNITVGGIIPGAGVVA-----RGF-----RGLOQVRVS-----DTPEGVNSLDPSHGE 1783
Db 2765 -----AGKAAVYGVTRTEFYVADYQKSLDIDRLRGLKPLPPANNG--TQWQ 2811
QY 1784 SI-----NVEQCGSLPDPDQSNPCPANSYCSNDMDYSQSC-----DPGYGDNCTNYCDL 1834
Db 2812 ATMARNLKQGPNSKPCSNVTCPPDFECVDLMNVYECTCGGRIMSPDSKGMORNECLD 2871
QY 1835 NPCEHQSVCIR-KPSAPHGTCCECPNYLGPYCE 1867
Db 2872 MPCMGAGATCINLEPRL--RYRCICPDGFMGENCE 2903

RESULT 14
ID PAT2_HUMAN STANDARD; PRT; 4349 AA.
AC Q9NYQ8; 075091; Q9NSR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (hPatz) (Multiple epidermal growth
DE factor-like domains 1).
GN PAT2 OR MEGF1 OR CDH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINB=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
RT feature of protocadherin genes."
RT Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 3777-4349 FROM N.A.
RX MEDLINB=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening."
RT Genomics 51:27-34(1998).
RN [3]
RP SEQUENCE OF 4142-4349 FROM N.A.
RX TISSUE=Brain;
RA Pouscka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 32 cadherin domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AF231022; AAF61928.1;
CC EMBL; AB011535; BAA32463.1;
CC EMBL; AL157443; CAB5663.1;
CC PIR; T46927; T46927.
CC Genew; HGNC:3596; PAT2.
CC MIM; 604269;
CC HSSP; P00740; IEDM.
CC GO; GO:0005509; F:calcium ion binding activity; NAS.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.

```



DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF000028; cadherin; 33.  
 DR Pfam; PF00008; EGF; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 31.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00282; Lame; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 14.  
 DR PROSITE; PS00268; CADHERIN\_2; 32.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 1.  
 DR Transmembrane; Glycoprotein; Repeat; EGF-like domain; signal.  
 FT SIGNAL 1 18  
 FT CHAIN 1 4349  
 FT DOMAIN 19 4048  
 FT TRANSMEM 4049 4069  
 FT DOMAIN 4070 4349  
 FT DOMAIN 34 148  
 FT DOMAIN 149 256  
 FT DOMAIN 363 458  
 FT DOMAIN 459 564  
 FT DOMAIN 565 669  
 FT DOMAIN 716 820  
 FT DOMAIN 821 925  
 FT DOMAIN 926 1032  
 FT DOMAIN 1033 1137  
 FT DOMAIN 1138 1242  
 FT DOMAIN 1243 1346  
 FT DOMAIN 1350 1448  
 FT DOMAIN 1449 1555  
 FT DOMAIN 1556 1660  
 FT DOMAIN 1661 1758  
 FT DOMAIN 1759 1872  
 FT DOMAIN 1969 2070  
 FT DOMAIN 2071 2171  
 FT DOMAIN 2172 2272  
 FT DOMAIN 2273 2379  
 FT DOMAIN 2380 2481  
 FT DOMAIN 2482 2585  
 FT DOMAIN 2586 2691  
 FT DOMAIN 2692 2797  
 FT DOMAIN 2798 2906  
 FT DOMAIN 2907 3011  
 FT DOMAIN 3012 3113  
 FT DOMAIN 3114 3218  
 FT DOMAIN 3219 3321  
 FT DOMAIN 3322 3426  
 FT DOMAIN 3427 3531  
 FT DOMAIN 3532 3642  
 FT DOMAIN 3643 3773  
 FT DOMAIN 3773 3944  
 FT DOMAIN 3947 3984  
 FT DOMAIN 3986 4022  
 FT DOMAIN 3297 3302  
 FT DISULFID 3951 3962  
 FT DISULFID 3956 3972  
 FT DISULFID 3974 3983  
 FT DISULFID 3990 4001  
 FT DISULFID 3995 4010  
 FT DISULFID 4012 4021  
 FT CARBOHYD 39 39  
 FT CARBOHYD 210 210  
 FT CARBOHYD 280 280  
 FT CARBOHYD 330 330  
 FT CARBOHYD 459 459  
 FT CARBOHYD 568 568  
 FT CARBOHYD 627 627  
 FT CARBOHYD 655 655  
 FT CARBOHYD 789 789  
 FT CARBOHYD 966 966  
 FT CARBOHYD 1175 1175  
 FT CARBOHYD 1303 1303  
 FT CARBOHYD 1383 1383

FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1904 1904 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1998 1998 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2007 2007 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2183 2183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2325 2325 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2368 2368 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2387 2387 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2430 2430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2470 2470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2547 2547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2597 2597 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2654 2654 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3125 3125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3276 3276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3310 3310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3430 3430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3471 3471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3601 3601 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3772 3772 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3813 3813 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3840 3840 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3873 3873 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3989 3989 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 4117 4117 P -> L (IN REF. 2).  
 FT CONFLICT 4160 4160 E -> G (IN REF. 2).  
 SQ SEQUENCE 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;  
 Query Match 8.5%; Score 1320.5; DB 1; Length 4349;  
 Best Local Similarity 26.1%; Pred. No. 5.8e-55;  
 Matches 478; Conservative 251; Mismatches 724; Indels 377; Gaps 65;  
 QY 175 VNT-----APQPPSYQATVPENQACTPVASLRAPDPSGRRLTYTMDALPDSRS 228  
 DB 2469 IYTNANKKSPFOOHLYEABLAENAMGTKYDILADKDSGPYCTIYT--IINKXA 2525  
 QY 229 NQFSSDPVTGAVTAAEELDRSTKTHV--FVTAQDHGMPPRSALATLTIVTDTNDH 286  
 DB 2526 SEKFSINP-NGQATLQKIDRNSSTERYALTYVMARDG-GRVAFYKILITDENDP 2582  
 QY 287 PVFQOEYKESLRNLENGYEVLTVAATDGPAPNANILYRLBSGSGSP---SEVPEI 342  
 DB 2583 PEFKASEYVTSIQSNVSKOSPVIQVLAADADGQADAVTVSV-----NPEDLVMDVIEI 2636  
 QY 343 DRSGYIRKRGVNDREKESYQLTVEASDQNDPGPRSTTAAPVFSVEDNDNAPQFSK 402  
 DB 2637 NPVTGVKVKDILVGLNQTLDPFITKAQDPG--PPHWSLVVRLQVVPKVSLEPFSER 2694  
 QY 403 RYVQVREDVTEGAP---VLRTASDRDKGSNAVVHYSIMSG-----NARQGFYDAQNG 454  
 DB 2695 LITFSAPEDLPBGSEIGIVKAAVADQ-----PYISLVKGTTPESNKQGFSLDPDTG 2747  
 QY 455 ADDVSPLDYETTKETTLRVARQDGRPLSNVSGL---VTVOVLINDNAPITVSTPF 510  
 DB 2748 VIKVRKPMDSHTKIXQIDVMAH-----CLQNTDVLSVSNVIQGVADVNDKPVFEDAPY 2802  
 QY 511 QATVLESVVLGVLTVAQALDADAGDNARLETRL-AGVG--HDPPTINNGTGWISVA 566  
 DB 2803 KAVLTENMEVGVSVIQTVAIDDTGRDGVSRLLSADPDSNVHER-FALDSBSGWIITLQ 2861  
 QY 567 ELDRREVDVDSGVGEARPDG--TPALTASASVTVLADVNDNPTPTOPEYTRLNEDAV 625  
 DB 2862 EIDDCICQYTHHVAHYDHGQITQLSSQALVSVITDERDNPAPRASEYRGSVENSESP 2921  
 QY 626 GTSVTVSAVDND--AHSVITYQITSGTNRNFSITSGGGGLVSLALPLDYKEROY 682  
 DB 2922 GELVATLKLTDADISEQRQVTCYITBGPPLQFGISQGVDEWRISSRKTLDRHTAKYL 2981  
 QY 683 LAVTASDGRQOTADIVNAVTDANTRHPVPSGSHYTVANVNERPAGTVVVISADEPDG 742  
 DB 2982 LKVTASDGRFQASVVEIIFVLVDVNDNSPQSQQLTYGKVRHEDVFGHFLIKVASATDLDTD 3041

```

QY 743 ENARTYEM-DSIFOPRIDATGAVTTOAEIDYEDOVSYTATATARDNGIPKSDTTYL 801
D 3042 TNAQITYSIHGPGAHFCLDHTGELTTLTALDERKOVPLVAKATDGC--GSSCADT 3099
QY 802 EILVNDVNDNAPQFLRDSYQGSVEYEDVPFTSVIQAISATDRDSCGLNEVFYTPQGGDGD 861
D 3100 TLHVEDVNDNAPRFPSSHCAVAFDNTVTKTPVAVFARPDQGANQVYSLP--DSAE 3157
QY 862 GDFIVESGIVRTIARLDRENAQVYLAAYADKGM--PARTMEYTVYLDVNDNPPV 920
D 3158 GHFSIDATGVIRLEKPLQVPRQAPLELTVRASDGTPEIPSTIGTIVSVVGGEDYLPV 3217
QY 921 FEODEFVFEENSPIGLAVARV--TATDP--EGTNAQIMYOIVEGNIPEVQOLDIFSGEL 978
D 3218 FLNTHESQVPEBDAPGTEVQIATLTPGAEKTC--YVVSNGEGRRLDARCTIL 3273
QY 979 TALVDLDEYEPYVLIQA--TSAPLVSRATVHRLDENDNPPVLANFEILFNNTYT 1035
D 3274 YVNASLDEPSTPKYPLSIECSRKSSSSLDVTVVNTVNEHRP--QPPQDPYST 3328
QY 1036 N-RSSSPFGAIGRPAHDP--ISDSLTSPERGENEISVILNASTGELSRALDNMR 1092
D 3329 RVLNHALGVILTVSATDEDEGPLNSDITTSIIGNQGHFTIHPKKGEQVAKALDREQ 3388
QY 1093 PLEAIMSVLSDG--VHSYTAOCALRVITITBEMLTHTSITLRLDMSPERFSLPILGL 1148
D 3389 ASSYSIKARATDSQCPRIHEDT-DIAIVADVND-----NPRFQOLNTST 3443
QY 1149 FIQAVATLATPPHVVVFNQRTDADGAILNVLNLSVGPPGCGPP-- 1199
D 3434 TVQ-----ENSPISGKVLQILIS--DPDSEPNPYSFRITKGN 3471
QY 1200 -----LSEDIQERLYNRLSLTASIQORVLPNDNCLREPCENMRCVYL 1247
D 3472 GSARVTPDGMVLVTRBEGSLRBAQEWYQIQASDSCIPPLSLTSVR-----VAVT 3522
QY 1248 RPDSSAP-----FIASSVLPRIHPVGLRCPGPGTGDYCETEVD--LCYSRP 1296
D 3523 EQSHAPALPLRIFITVGEDEBQ-----GAWGKIHARDRDQDILTYS-- 3567
QY 1297 CGPHRCRSRGGYTCLCRDGTGHEHCYSASRGCTPGVCNKGCTCVNLLVGGFKDCP 1356
D 3568 -----LAEBETLGHFVGVAPDG-- 3586
QY 1357 SGDFEKPCQVTTSPHASFITPRGAFORHFTLALSFATKRDGLLTVNGRNEKHF 1416
D 3587 -----IIAAGL-PRGHT--SFNTVSDGTFTTTAGVAVYVM 3621
QY 1417 VALFVIOQVOLTFESAGESTTVSPVPGVSDQGMHTVQKYYNKPLAQGTGLPOGFSR 1476
D 3622 VGOEALQOAMMMGF-----YQLTREBELVSDHMRQLQFLSHKIDIKRANHLSLQ 3672
QY 1477 OKVAVVTYDGDCTGVALRFGSVLANSYCAAOCTGGSKSLDLTGPLLGGVPDLPBSFP 1536
D 3673 PAEAVAGVD--VLAVF-----BGHSSTFEYFQELAS-IITSAKMEHSV 3715
QY 1537 VRMRQFVCGMKNLQVDSHIDMADFIANGTVPGCPAKVNCDSNTGNGTGVNQMDAF 1596
D 3716 VQMR-----SAMPVPCGPPCCOGICN-----TV 3741
QY 1597 SCECELGFGKSCAQMANTPOH-----FIASSLVAMHGLSLPISQFWYSLT 1642
D 3742 HLDPRVGPYSTARSLITPRHRLQSCSCNGTATRFGGQYVYRA--PAARNWHIF 3798
QY 1643 MFRTRQADGVLLQATTRGSTITLQREGHMLSTVEGTGLDASSIRLEPGANGDMH-- 1700
D 3799 YLKTLPQALIL--FTNETASVSLKASGVQLEYNCLGSGYVGNLSSQ-RHVNHEHWSI 3855
QY 1701 -----HAQLALASGPGHAILSPDYGOORABENGLPRLHGLHSNITVGG--IPPA 1751
D 3856 LVEBMDASIRLNVDSMNTSLV-----PENCNG-LAPREH-----LLLGILLHSS 3902

```

```

QY 1752 GGVAQGFRCGLQGVRSDFREGVNSLDPHSG-----ESINVEQCSLPDPCDSNCPRANS 1806
D 3903 SNVSGEGLDVAVYNE--EALDLAPKTVAGLLETQALTOCCLSHSDYCSQNTCLNG 3960
QY 1807 YCSNDMD--SYSCSDPGYGDNC--TNVCDLNPCHQSVCYTRKPSAPHGYTCBCPPN 1860
D 3961 KCS--WTHGAGYCKCPQPGFSGHRCGREGNCTFAFLCBGCTCI--LSPKASCNCFHP 4015
QY 1861 YLGPYCEYR-----IDQPCPPGWWG 1880
D 4016 YTDRCCEMARGCSGHCVLTPETIQRGDWG 4045

RESULT 15
FAT2 FAT2_RAT STANDARD; PRT; 4351 AA.
AC 088277;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (Multiple epidermal growth factor-like domains 1).
GN FAT2 OR MEGF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MBLIN=98360089; PubMed=9693030;
RA Nakajima M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
CC -1- SIMILARITY: Contains 33 cadherin domains.
CC -1- SIMILARITY: Contains 1 laminin G-like domain.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
D EMBL; AB011527; BA032458.1; .
D PIR; T00252; T00252.
D HSP; P08709; IBF9.
D InterPro; IPR002126; Cadherin.
D InterPro; IPR000742; EGF 2.
D InterPro; IPR006209; EGF-like.
D InterPro; IPR006210; IEGF.
D InterPro; IPR001791; Laminin G.
D Pfam; PF00028; cadherin; 33.
D Pfam; PF00008; EGF; 2.
D Pfam; PF00054; laminin G; 1.
D PRINTS; PR00205; CADHERIN.
D SMART; SM00112; CA; 33.
D SMART; SM00181; EGF; 2.
D SMART; SM00282; Lang; 1.
D PROSITE; PS00232; CADHERIN_1; 14.
D PROSITE; PS50268; CADHERIN_2; 33.
D PROSITE; PS50025; LAM G DOMAIN; 1.
D PROSITE; PS00022; EGF_1; 2.
D PROSITE; PS00186; EGF_2; 2.
D Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
D SIGNAL 18
D CHAIN 19 4351
D DOMAIN 19 4050
D TRANSMEM 4051 4071
D POTENTIAL 4072 4351
D CYTOPLASMIC (POTENTIAL).

```

FT	CARBOHYD	3774	3774	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3815	3815	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3842	3842	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3875	3875	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3906	3906	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3991	3991	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	4351 AA;	480644 NM;	C564E7F98BB388	CRC64;
Query Match 8.2%; Score 1275; DB 1; Length 4351;					
Best Local Similarity 24.0%; Pred. 12.9e-53;					
Matches 534; Conservative 271; Mismatches 830; Indels 588; Gaps 777;					
QY	179	POPOPSPYOATVPENOPAGTPPVASLRALIDPDBEAGRLERYTMDALFDSRSNQPSFLDPVT	238		
DB	2377	PKFPEBPQYEAUVNSLACGHLVLKVQALDPDIDGTSRLFY---	LILSGNQDRHSINSTS	2433	
QY	239	GAVTAAEILRETSYHVFRTYADDHGMPRSALATLTLITVTYDNDHDVPVEQOEYKSL	298		
DB	2434	GIEMFNLCCKQQLDSSVNLRVGASD-GVFR--AAVPAVINTNANKSPERQOVVYEAEL	2490		
QY	299	RENLEVEYELVTVAATGDAAPNANNILYRLLESGSGSPSEVEFLIDPSSGVRTRGVPDRE	358		
DB	2491	AENAKVGTKEVLEALIDDSGPGYGVDTYTIINKLAG--ERFFINPR-GQITTLQKLDRE	2548		
QY	359	EVESYQLTVEASDGRDPGPRSTTAAVFLSEVEDNDNAPOFSEKRYVQVREDTYPCAV	418		
DB	2547	--NSTERYIALIKWARBDGGKVAFCYKILITDENDNAPOFKASGYTYSISNVSRSPI	2604		
QY	419	LKVTASDRDKSNAVHYSYTS-----	440		
DB	2605	IQVLAADYADBEGRNADVTYSVDSTEDLABEILFVNPTGVVKEKSELVLEBRADVFNKA	2664		
QY	441	-----	440		
DB	2665	QDGPPEHMDSLVPRLOVPENEPLPRPSEPLTYSAPEDIPESESGISGVAKAAQDBI	2724		
QY	441	-----GNAGQFYLDACQTALDVSPHLDYETKEVYTLRYRA-----ODGGRPLSN	486		
DB	2725	YSLVQGTTPBSNSDDVFLDDQDTGLKVRKAMDHESITKMQIDLMACHPHEDTDLVSLVS	278		
QY	487	VSGLVTVQVLDINDNAIFVSTPPQATVLESVPLGYLVHQAIDADAGNAALEFRLA-	545		
DB	2785	VS-----IQVEBVMDNRVFEADPKYAKRLTEMPGSTVVIQVTANDDOTGSGQVYSRLSV	2840		
QY	546	--GVGDHPFPPTINNGTMSVAALDEEBVDYFSFGYEARDHG--TPALITASASVTVLD	602		
DB	2841	EPGSHIELRAVDSESGNITTLQELDCEQTQYTRFYVAFDHGQITQLSSGALVEVSTTD	2900		
QY	603	VNDNNPTFTQPEYVYRLNEDAAVGTSVVTVASAVDRD---HSVTTYOITSQNTNRFSIT	659		
DB	2901	ENDNPPFAEDBRXGSVENNEPELVATIKTLDADVSDQNRQVTCYITBBDPLGQPSIS	2966		
QY	660	SQSGGGLVSLALPLDYLERQYLVATASDGTRODTAQIVNVVTNANTHRPVGOSHVT	719		
DB	2961	QVDEMRISRSKTLDRHIAKYLIRATASDGKFOQASPVAVEFVVDINDNSPOCQLLYTG	3020		
QY	720	NVNEDRAGCTTVVLISATDEDTGNAATYPMF--DSLTPORLADRGATVQALEDYDQ	778		
DB	3021	KVRSDVTPGHFLKVSALIDVDMTNAQITYSLHGPGAQEFPLDHTBELTTLTVLDERK	3080		
QY	779	VSYLATALTAADNGIPQASDPTVYLEILVNDVNDNAPOFLRDSYQGSVEDEVPPFTSVQLIS	838		
DB	3081	DVNVLVAKAIDGG--GSGCAEATYLIHEVDNDAFRFPBHCDAVAVDNTTVKTPAVVF	3138		
DB	3139	ARDDPOCANNAQVYSILT--DSADGQFSDIDATSGVIRLEKPLQVAPASAVEITVAPASDLGT	3196		
QY	839	ATDRDSGLNGRVFTTFQGGDGDGDFIVESTSGIVTRLRLRDRENVAAQVYLRAVADKGM	898		
QY	899	P-PARTMETVTVTVLDVNDNPPVPEODEPQVFEVENSPIGLAVARV-TANDP-DEGNAQ	955		
DB	3197	PIPLSTLGTGVTVSGVEDLPLPLFNAHESHTQVDEDPADIMEVLAHLATLLTRPGSEKKG-	3254		
QY	956	IMVOYIGVAGINPEVFOQLDIFSGEGLTALVDLDYEDRPEVLAIQA---TSAPLVSRATVHR	1012		

```

Db 3255 --YITGNEQKRLAHITLIVNGSLDEINPKYFLSIECRKSSSISDVITVIN 3312
Qy 1013 LLDENDNP-----VLGNPEILFNNYVTRSSSPFGAIGRPAHPD--ISDSLVSFER 1066
Db 3313 VTDVNEHFRTHLTVRVLENV-----GDVILTVSADDDGPNVSAITVYLVG 3364
Qy 1067 GNEILSLVLNASTGBELKSLRLLDNNRPLEAIMSVLSDG---VHSVTAQCALRVITTD 1122
Db 3365 GNGIGHFTINPKKGLQVAKALDWEQTPSYSLRLRATDSGQPLHEPT-EVAEVDVND 3423
Qy 1123 EMLTHSTLRLEDMSPERFLSLGLPIQAVATLAPPDHVVVFNQORDADAGHIILN 1182
Db 3424 -----NPRRFQ-----LMYSTSVQNSPIG-IKV 3447
Qy 1183 VLSVGOBPGPGGPPFL-----PSEDLQERLYINRSLITAIQAQVLPFDNICLEBPC 1237
Db 3448 LQILIDBDSBQNGPFRITBGNISGVFRVTDGRLVTAASL-----KKAR 3496
Qy 1238 ENYRCVSVLRFSSAPFIASSVLFREIHVGLRCCRP-----G 1279
Db 3497 EMYOLHIEV--SDSGLPPLSSSTLVRVQVTE---QSRYPSTLPLRISITKGEERFQGG 3550
Qy 1280 FTGDYCEBYD---LCYSRPGHGRGRSREGYTCLCRDGYTGHCEVSAASGRCTPG 1335
Db 3551 MIGIHAIDRDPDPTLTS-----LEOEGGLDRYFTVGSADGK--- 3588
Qy 1336 VCKKGGTCVNLVGGFKCDPSGDEKPYCOVTRSPRASHFITFRGL-RORFHTLALS 1394
Db 3589 -----IASQGLPHGRYGFNVTVS 3607
Qy 1395 FAYTERDGLLYNGRFEKHDVALLVIOBOVLTFSAGSSTTVSPFVPGVSDGQWHT 1454
Db 3608 ---DGTFTTGGVHVHVMMEBEVQAVWLGF-----HOLTPELVSDBWRN 3652
Qy 1455 VOLKYKPBLLGQGLPGSPSEQKAVVTVNDCDITVALFRGSLVGNVSCAOGTOGSK 1514
Db 3653 LQRLSLNLDVKRANIHILASIQPAEVTAGVD---VLVLF-----ERHSGTS 3695
Qy 1515 KSLDLTGFLILGVPDLPESEFVVRKQFVGCMRLQVDSRHIDMADFIANNGTVPGCCPAK 1574
Db 3696 YDLOELASAIHNSVREIHSVIGIR-----SALPVVPCQ 3730
Qy 1575 KNVCDNSTCHNGGTCVNQMDAPSCCEPLGFGGKSCAQEMANPOHFLGSSLVAMHGLSPI 1634
Db 3731 GQSCODQTCQ-----ETVSLERVRGVSSTARLSILTPRHILGRN-CSCNGTTLRF 3780
Qy 1635 S-----QPMYLSLFRTRQADGVLLQAITRGRSTITTOREG-HVMLSVEGT 1680
Db 3781 SGQSVQYRPLEADQWQIHFTYKTLQPMALLM--FTNETASISLKLANGFSLHYHCPG- 3837
Qy 1681 GLQASLLEFRANDGDMMHQAQLALGASGFGHAILSPDYGOQRAEGNLGPRILGLHLS 1740
Db 3838 GFYGNLSRRYP--VNDGGMH--SMLEBRDTSVHLVDI---TDNLSLVIPECOGLRTE 3890
Qy 1741 -NITVGGI--DPAAGVARGRGCLQGYRVG-DTPEGVNSLDPSHG--ESINVEQCCSLP 1794
Db 3891 RQLLIGLIVPNSPSSNVSLGFEGLDVAVNGERLELLGRKKMEGRLETWALSQCCWPG 3950
Qy 1795 DPCDSNPPRANSYCSNDMS--YSGSCDYGYYGDNQ---TNVCDLNPCENHOSVCTRKPAP 1850
Db 3951 TACGSPCLNGSGCPALGSGYLKCPFPFSGRNCLEGRNCTSAPOEGGTCV---SSP 4007
Qy 1851 HGYTECEPPNYLGPYCETR-----IDQPCPRGMWGH-----PTCG 1885
Db 4008 EGTSCNCFHPYTGRCCEMBARGCSGHCILITPEIKRGMGOQEBLVITVALPLVITATVG 4067
Qy 1886 -----PCNCVSKGDPDPCAKTS-----GECHCKENHYRP 1915
Db 4068 LLLYCRRRKSHKPYTME---DPLILARSIGVDTQASPAIELDPLNTSSCN---NLNQP 4119
Qy 1916 PGSPTCLCDCVPTGSLSR---YCDPBDGQPC---KPGVIGRQCDRCNPFABVTTN 1967

```

```

Db 4120 EPRKTSVPELVTGSSSKQRPWVCVPPRLPPAAVSSHG-----HEPIIKRTWS 4170
Qy 1968 GCEVNTDSCPRATEAGIM-----W-----PRTFRGLPAAPCPKGS 2003
Db 4171 GEELVIPS-----GAAVPEPTYSRKGMEXEPHETMOGILPSPRHHVG-PAVMPDPTGL 4224
Qy 2004 FG--TAVRHCDERGMPLPPLFN-----C-----TSITF----- 2030
Db 4225 YGCFPEPLENKRALPPRYSNQNLDMPPRPPSPREHLAPCLNEVTAISYHSQFR 4284
Qy 2031 -----SELKGFARLQNR-----BSGLDSGRSQQLALLRNATQHTAGYFG 2071
Db 4285 QGGGGBCLAEGGYKGVSMRLSRAGPSYADCEVNGGATGRSQPRA-----PNYEG 4335
Qy 2072 SDV 2074
Db 4336 SDM 4338

```

Search completed: February 11, 2004, 15:55:45  
 Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 15:52:09 / Search time 76 Seconds  
(without alignments)  
9924.839 Million cell updates/sec

Title: US-09-916-849A-3

Perfect score: 15545  
Sequence: 1 MKSPATGVPLPTPPPLLL.....ACTVDESSGSEPLFFNPLH 2923

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_ricent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5065.5	32.6	1766	13 Q8AW45	Q8AW45 brachydanio
2	2491	16.0	2610	5 Q19482	Q19482 caenorhabdi
3	1538	9.9	2909	5 Q8MY78	Q8MY78 ascerina pe
4	1520	9.8	4589	11 Q9WU10	Q9WU10 ratius norv
5	1519	9.8	4610	13 Q8AX87	Q8AX87 brachydanio
6	1508.5	9.7	2809	5 O61230	O61230 lyechinus
7	1496	9.6	4587	11 Q9QX3	Q9QX3 mus musculu
8	1413	9.1	4555	11 Q8R508	Q8R508 ratius norv
9	1164.5	7.5	2920	5 Q967F4	Q967F4 caenorhabdi
10	1108.5	7.1	1959	5 Q8MY75	Q8MY75 ptychodera
11	1092	7.0	4307	5 Q19319	Q19319 caenorhabdi
12	1081.5	7.0	1061	4 Q8NSB3	Q8NSB3 homo sapien
13	906	5.8	1881	5 Q8T5J8	Q8T5J8 anopheles g
14	905.5	5.8	803	4 Q96JL3	Q96JL3 homo sapien
15	905	5.8	1820	5 Q9VVG0	Q9VVG0 drosophila
16	905	5.8	1820	5 Q8IGX4	Q8IGX4 drosophila

17	862.5	5.5	2005	5 Q9VFH5	Q9VFH5 drosophila
18	853	5.5	1060	4 Q8IUP2	Q8IUP2 homo sapien
19	850	5.5	1037	4 Q9H4E1	Q9H4E1 homo sapien
20	850	5.5	1037	4 Q8MY34	Q8MY34 homo sapien
21	850	5.5	1048	4 Q9BZ49	Q9BZ49 homo sapien
22	850	5.5	1340	4 Q9BZ48	Q9BZ48 homo sapien
23	848	5.5	1038	11 Q8CFX3	Q8CFX3 mus musculu
24	847	5.4	1021	4 Q9P2X5	Q9P2X5 mus musculu
25	847	5.4	1025	4 Q9H4E0	Q9H4E0 homo sapien
26	847	5.4	1337	4 Q9BZ46	Q9BZ46 homo sapien
27	847	5.4	1347	4 Q9BZ47	Q9BZ47 homo sapien
28	833	5.4	1310	4 Q96RV4	Q96RV4 homo sapien
29	833	5.4	1347	4 Q96RV0	Q96RV0 homo sapien
30	831.5	5.3	965	11 Q8QZT6	Q8QZT6 mus musculu
31	814.5	5.2	961	4 Q8ND89	Q8ND89 homo sapien
32	809.5	5.2	1032	6 Q9SKD8	Q9SKD8 macaca fasc
33	809.5	5.2	1203	4 Q9HCE6	Q9HCE6 homo sapien
34	786.5	5.1	1016	13 Q9QZT2	Q9QZT2 xenopus lae
35	780.5	5.0	1035	13 Q57537	Q57537 xenopus lae
36	768.5	4.9	1069	11 Q8B185	Q8B185 mus musculu
37	753	4.8	1563	5 Q9VAF5	Q9VAF5 drosophila
38	739	4.8	733	6 Q9SKX0	Q9SKX0 macaca fasc
39	738	4.7	1084	4 Q8N3K7	Q8N3K7 homo sapien
40	735.5	4.7	1469	4 Q9HAR3	Q9HAR3 homo sapien
41	735.5	4.7	1474	4 Q94910	Q94910 homo sapien
42	733.5	4.7	1030	13 Q9QZT4	Q9QZT4 gallus gall
43	729.5	4.7	1467	6 Q97830	Q97830 bos taurus
44	729.5	4.7	1472	6 Q97831	Q97831 bos taurus
45	728	4.7	1515	11 Q88917	Q88917 ratius norv

## ALIGNMENTS

## RESULT 1

ID	Q8AW45	PRELIMINARY;	PRT;	1766 AA.
AC	Q8AW45;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	SI:BZ6108.1 (Novel protein similar to cadherin, KDP IAG seven-pass G-type receptor (CRLSR)) (Fragment).			
GN	SI:BZ6108.1.			
OS	Brachydanio rerio (zebrafish) (Danio rerio).			
OC	Bryozoa; Metazoa; Chordata; Vertebrata; Osteichthyes; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.			
OC	NCBI_TaxID=7955;			
OX	NCBI_TaxID=7955;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Babbage A.			
RL	Submitted (DRC-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, AL645790; CDS8738.1; -.			
FT	NON_TER 1 1			
SQ	SEQUENCE 1766 AA; 197485 MW; 80CAAA18F5A497F CRC64;			

Query Match	32.6%; Score 5065.5; DB 13; Length 1766;
Best Local Similarity	52.3%; Pred. No. 8.5e-312;
Matches 976; Conservative 279; Mismatches 431; Indels 181; Gaps 37;	
QY	1105 GVHSVTAQALRVITIDEMTHSITRLBDMSEPLSLGLFIQVNAATLTPPDHV 1164
DB	1 GHQVTAFCRLRVITIDEMTHSITRLBDMSEPLSLGLFIQVNAATLTPPDHV 60
QY	1165 VFWVQDPTDAPGHHIIVSVSQPPGPGGPP--FLPSDLOERLYIKNSLITLTAISQ 1222
DB	61 FIFVQNDTV--SGNINLVFSALP---GGVDRYFPESELOEQLYIKNTLLQELISSQ 115
QY	1223 RVLPFDNIIKREPCENYKRCVSLRFPDSSAPFASSTVFRPIHPGGLRCRCPPPTG 1282
DB	116 NVLPFDNIIKREPCENYKRCVSLRFPDSSAPFASSTVFRPIHPGGLRCRC----- 169

QY 1283 DYCETEVLDLCYRBPQPHGRCSREGVYCLCRDGYTGEHCESVARSGRCTPGVCKXNGT 1342  
 DB 170 -----ENCEVDSRGRCVPGCKNGE 191  
 QY 1343 CYNLLVGGFKDCDSCGDEPEKYCOVTTTRSPPAHSFTRGRLRQNFHTLALSPYTKERDG 1402  
 DB 192 CYNLLVGGFKDCDSCGDEPEKYCOVTTTRSPPAHSFTRGRLRQNFHTLALSPYTKERDG 251  
 QY 1403 LLLYNGRPNRKHDPVLALEVIQEOYOLFRSAGESITTVSPVPGVSDGQMTVYOLKYYNK 1462  
 DB 252 LLLYNGRPNRKHDPVLALEVIQEOYOLFRSAGESITTVSPVPGVSDGQMTVYOLKYYNK 311  
 QY 1463 PLLGQTLPGPSBQKXAVVTVDCDGTVALRFGSVLGNVSCAAQGTQGSKSLDITGP 1522  
 DB 312 PKISHLGLPGPSBQKXAVVTVDCDGTVALRFGSVLGNVSCAAQGTQGSKSLDITGP 371  
 QY 1523 LLLGCVDPDLPSPFVRKQPFVGCNRNLOVDSRHIDMADFIANNGTVPCCPAKXVCDSNT 1582  
 DB 372 LLLGCVDPDLPSPFVRKQPFVGCNRNLOVDSRHIDMADFIANNGTVPCCPAKXVCDSNT 414  
 QY 1583 CHNGGTQVQNDARSCECPFGGKSCAQEMANPOHFLGSSLVAMHGLSLPISQPTVLSL 1642  
 DB 415 CQHAQENCKNNTFCCEPBGKXKCDQDMPKOHFGHAMMLNDDMTIAVFWTIAL 474  
 QY 1643 MFRTRQAD--GVLLQAITRGRSTITTLQRBGVMLSVBGTGLQASSLRLEPGRANDGDWH 1700  
 DB 475 MFRTRQAD--GVLLQAITRGRSTITTLQRBGVMLSVBGTGLQASSLRLEPGRANDGDWH 519  
 QY 1701 HAQALASAGS-----FGAHLISPDVGOQABEGNLGRLGLHLSNITVGGITGPAGV 1754  
 DB 520 HLLVLELSSKDGKDKTKMAQVLLDYDMFKSVE--IGNELPGLTKKSPFIOGLOQODIV 577  
 QY 1755 ARGRCGLQVAVSDTEGVNSLDPSSHGESINVEQSGSLPDCPSNCPANSYCSNMDS 1814  
 DB 578 QQGGKGMQGLARMETATSTANIMHHAQKTRVEDGCDMSANCALINCPKSOCT--DE 634  
 QY 1815 YSCSGDPGYGDNCTNVCNLPCEHQSVCTRKSPAPHGTYCECPNLYGPYCETRIDQPC 1874  
 DB 635 HTCTCDGPGFGRDQVADACHLNPCEHLSTCYRKPFSSSHGTYCEGQDYGYQCEKNEKPC 694  
 QY 1875 PRGMWGHPTQGPCNCDVSKGFPDCKNTSGRCHCKENHPRPGSPFTCLDCYCTGSLR 1934  
 DB 695 PRGMWGHPTQGPCNCDVSKGFPDCKNTSGRCHCKENHPRPGSPFTCLDCYCTGSLR 754  
 QY 1935 VCDPEBQCPCKPVIGRQCDRCNPFALVTTCGCVNYDSCPAIBAGIWMPTRRGLP 1994  
 DB 755 TCDBETQCPCKKAVIGRQCDRCNPFALVTTCGCVNYDSCPAIBAGIWMPTRRGLP 814  
 QY 1995 AAAPCPKSGFPTAVRHCDEHKGWLPNULFNCTSTPSELKGFARLQRBESGLDGRSQ 2054  
 DB 815 VAMNCPKSGISTAVRHCDEHKGWLPNULFNCTSTPSELKGFARLQRBESGLDGRSQ 874  
 QY 2055 LALLRYAHTOHTAAGFSDVAVAYQATRLLAHSTQGRGLSTQVHPTRENLARTGSA 2114  
 DB 875 IARLNAHTOHTAAGFSDVAVAYQATRLLAHSTQGRGLSTQVHPTRENLARTGSA 934  
 QY 2115 LLDATANRAMELIQTEGGTAMLIQHYEAYASALQNMKRYLSPFTIVENIYISVRL 2174  
 DB 935 ILDSNKEHMOIQRTBGTANLKHPEYANTLAQNMKRYLSPFTIVENIYISVRL 994  
 QY 2175 ----DKGNFAGAKLPRYEALRGEPDLETTVILPESVRETTPVVPAGEAQEPBEL 2230  
 DB 995 VTYPD-----ATKYPNFQTDQECFKDKSSVLPPESEFKSKHEDBEQINSDEEEKQ 1049  
 QY 2231 A-RBQRHPELSQGBAVASVIYRTIAGLPHNTPDPRKSLRVKRPINPVVISVHD 2289  
 DB 1050 ATKPKRKRVTITPPPPASVITIKTLGQFLEHNDPDRKSLRVKRPINPVVISVHD 1109  
 QY 2290 DEBLPRALDPVTVQFLLETERKIKCYFMNHSILVSGTGMASRGCEVVRNESHV 2349  
 DB 1110 EGQPLHPIPLERITLIDVNLLETBERTKVCVFMNHSITVGAGAMSKGCDIISRNTHTI 1169  
 QY 2350 SCQCNHMTSPAVLMDVSRRENGEILPKTLITYVALGVTLAALLTPFPFLTLRLIRSNQ 2409

DB 1170 SCQCNHMTSPAVLMDVSRRENGEILPKTLITYVALGVTLAALLTPFPFLTLRLIRSNQ 1229  
 QY 2410 GIRRNITAGLQVFLGINDADIPACTVAILLHFLYCTFSMALLBALHYALT 2469  
 DB 1230 SHKNVVALFESLEVLGINDADIPACTVAILLHFLYCTFSMALLBALHYALT 1289  
 QY 2470 EVRDVNTGPMRPFYVNGVGPVATITGLAVGLDEBGVNPFCVLSYDITLINSFAGVAF 2529  
 DB 1290 EVRNINHGMRFTYALGKGPALITGLAVGLDQSGNPNFCVLSYDITLINSFAGVAF 1349  
 QY 2530 AVMSVFLYTLAARASCAARQGFBEKKGPVSLQPSFAVLLLSATWLALLSVNSDTLL 2589  
 DB 1350 VLVNIVIFWMAKASCGRRQSEYSGVTPALRMAFLILLISATWLGLMAVNSDVT 1409  
 QY 2590 FHTLPATCNCIOGPPFLSVLVSKEVRALK-IASRKRSPPPALITST-LTSSYCP 2647  
 DB 1410 FHTLPATCNCIOGPPFLSVLVSKEVRALK-IASRKRSPPPALITST-LTSSYCP 1469  
 QY 2648 SPY-ADGRLYQ-PYGPAGSLHSTSRSGKSPSYIFPFLRBSGALNP--GQPPGLGD- 2701  
 DB 1470 NTMEDGALFRALIGESTVLSSTVSAKSHSYFTYTRDELQKPSVGNKAGITDI 1529  
 QY 2702 PGSLFLEGQDQHDPTDSDSLSLBDQGSYASTHSSDSEEBEEREAAPGBOGW 2761  
 DB 1530 DGSILFRNGTVAD--DSDSDSELV-DEHSSYASHSSDSEEDIDIMQPK-----W 1577  
 QY 2762 DSIHGGAERLPHHSTPK-----DGGPGKAPWPDFTGTAKSSGNGAPERLRENG 2816  
 DB 1578 NN-----EKOPHLSPTKTEVDVTVSNHVKPWPPEAMTASDSEBDGA--ERLVEYK 1629  
 QY 2817 A--LSREGSLG-----PLP-----GSSAQP--HKGLKKKCL--PTISERK--S 2852  
 DB 1630 VNTMLQENKMLNIGSTQDKQPTPSIQANSNHOEGKGLIKNTITTYPTIDCMK 1689  
 QY 2853 LIRLPLEQC--TGSSRGSS--ASEGSRG--PPRP--PPROGLQEOQNCVPMPIAMS 2901  
 DB 1690 KIREKLSVYMPPTISSRAVITSNDGANGHNVILKPPRPVPR--EHLANG--MAN 1742  
 QY 2902 IKAGTV 2908  
 DB 1743 LKSAV 1749  
 RESULT 2  
 ID 019482 PRELIMINARY; PRT: 2610 AA.  
 AC 019482; Q23218;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE P15B9.7 protein.  
 GN P15B9.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Baynes C.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RP SEQUENCE FROM N.A.  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Riken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sprat J., Mohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL elegans.";  
 RN Nature 368:32-38 (1994).  
 RP [3]  
 RA SEQUENCE FROM N.A.  
 RA Percy C.;  
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: Z78018; CAB01449.1; -.  
 DR EMBL: Z78013; CAB01449.1; JOINED.  
 DR EMBL: Z78013; CAB01427.1; -.  
 DR EMBL: Z78018; CAB01427.1; JOINED.  
 DR HSP: P15116; INCI.  
 DR WormPep; F15B9.7; CE09412.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006209; EGF\_Like.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR001791; laminin\_G.  
 DR InterPro; IPR000203; PKD\_cye\_rich.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF\_2.  
 DR Pfam; PF01825; GPs; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 8.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00303; GPs; 1.  
 DR SMART; SM00303; GPs; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 6.  
 DR PROSITE; PS0268; CADHERIN\_2; 8.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS50227; G\_PROTEIN\_RECPT\_P2\_3; 1.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 2.  
 DR KMW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KM Glycoprotein; Repeat; Transmembrane; EGF-like domain;  
 SQ SEQUENCE 2610 AA; 290075 MW; 24184B4C836805C4 CRC64;

Query Match 16.0%; Score 2491; DB 5; Length 2610;  
 Best Local Similarity 27.4%; Pred. No. 4e-146;  
 Matches 768; Conservative 438; Mismatches 1205; Indels 394; Gaps 89;

QY 51 CAPMGLCPSSASNLMLYTSR--CRDAGTELGHVPHDGLRWCPSESAHILPPAPE 108  
 DB 39 CRKCA--VPSNSNVIMLPASRPCLHPGPPII--HMPDISDILA-----CPVGLPD 86  
 QY 109 GCMWCRLLGIGHLSPQGLTLPEHPC-LKAP-----RLRQSCSLADAPGL 156  
 DB 87 SVHSQSLSL-----LRGEG-LTLTKERICFPDGFIDFHYDVCDDGLYKSKMTIGHSIAS 140  
 QY 157 RAGERSPEESIGGRKRKNVNTAPQPPSYQATVPENQAPGTVPASLRATIDPGRAGRL 216  
 DB 141 KKLLEFRRTKRWARRRNPDAVHFOEKVYKELPEBTPLETIIVASVKA---SHASQPL 197  
 QY 217 EYTMALPDSRNOFSLDPVTGAVTTAEELDEKTSVHFVRYTAODHGNRRSALATLT 276  
 DB 198 YYSVAVPODSRSQNLFTLDTMSGEIRLAKMDREVDLKHILKTYAERVDPITISASTTV 257  
 QY 277 ILVTDNDHDPVEQGEYKESLRLENLEVEYVLTVAATGDAPNNILYRLLEGGSGSP 336  
 DB 258 VHVLDVODNSPIREKSYFGEIREDAPIGTVLVSFARDDSGENGIEYSLGSGNG--- 314  
 QY 337 SSVFEIDPRSGVIRTRGPVREBEVSYQLTVBASDGRDPGPRSTTAAVFLSYEDNDNA 396  
 DB 315 KNLAINAKSGVIGTAPALDRFTLSLRDLVIASDKGP--KRESTAMWEITVVVDVNDNA 372  
 QY 397 POSSEKRYVQVEDVTGAPVLRVTAASDDKSNVAVHYSNGNARQGFYIDAQTGL 456  
 DB 397 POSSEKRYVQVEDVTGAPVLRVTAASDDKSNVAVHYSNGNARQGFYIDAQTGL 456

DB 373 PVFASDYNVTILENITTPAVIATVKATDEDEFTNGKHYVSSASSGIGLTIIDYSTGEV 432  
 QY 457 DVVSPLDVETTKETLARAADGGRPLSNVSGLVTVQVLDINDNAPIVSPFOATVLE 516  
 DB 433 TLREIRIDAKNS-PITAVIRAKDGAQPALSTVPL-TINVIDINDAPVILIAQKMITLEE 490  
 QY 517 SVPLGYLVHVOAIDADAGDNARLEYLAGVGHDPPTINNCGMIVSAEILDEEVDY 576  
 DB 491 NVAIGBEVGVVAIDEDSPNGIIRYSMEG---SDFIIDESGLIKTKLDRRTTAY 547  
 QY 577 SGPVEARDHGIPALPASASVTVLDVNDNFTFPQPEYTVRLNEDAAVGSVTVVSAVD 636  
 DB 548 SLKVTARMDGTPSLTSTTIAVVLKDINDNAPTFPKENYVITISEMPGSOITLKAVD 607  
 QY 637 RDAHSVITQITSGNTRNRSITSSGGGLVSLALPDVKLERO---YLAATASD-GTR 692  
 DB 608 NDEOKITRIEAD-REVSFLIDIGDQAI--LSVSEELRQDHKVAVEISATDQGL 663  
 QY 693 QDTAIVNAVTDANTHRPVQSSHYTVVNNEDRPAGTTVLIS----- 735  
 DB 664 QGRCVNVPIIDVNS-APYFNHPSVVKIPEHSPIGPVITLKVSFFSYFGYGLKHFV 722  
 QY 736 -----ATDEDTGENARITYFMEDSIPOFRIDADGAVTTQALDEYEDV 779  
 DB 723 SGNLRKKRIPNSANMHAEDHREDGNARIVYSI-DSSQFERIDPSSGDISVSSDLREDRA 781  
 QY 780 SYTLAITARDNGIPKSDPTVLEILVNDVNDAPQFLDSYCGSYVEDPPTSVYQISA 839  
 DB 782 TFSVIVLASDHAAPLANTSTQLEVLIDINDNSPQFTSSSVATISEDIPIVGTSPQVSA 841  
 QY 840 TDRDSGLNGRVFYTEQGGDGDGD--FVIESSTGIVRTLRLDRRENVQVYLAAYADK 897  
 DB 842 IDADIGNGVIVDYFLNESSSSPSIQFLDRISGTLRYSKLDKQFAVIVLPIARBRG 901  
 QY 896 MPAPATMEVTVTVLDVNDNPPVFEQDEDFVEENSPIGLAIVARVATADPEGTNAQIM 957  
 DB 902 TFSLSAASEITLTLDVNDNAPTFEQLSYDLIAENSFGSVTGTIVADADDEGNADIS 961  
 QY 956 YQIVGKNIPEVQQLDIFSGE-----LTLVLDYEDDRP-EYVLVIOATSAPIVSRATVY 1011  
 DB 962 FRIFGADAKLDIEDBQNGVVRILTRAEVDYAKANKPFPELOASGQSLSTVPVPI 1021  
 QY 1012 RLIDRNDNPPVLANEELFNNTVTRSSFPGALIGRAVHPDIDSILTSYFPERGNETLS 1071  
 DB 1022 HVSVDNNDKPKADPVLILMNRDYNQMAR---QIGFIPAPPDQNAITLEYLE---END 1074  
 QY 1072 LVYLAASGELKLSALDNNRPLBAIMSVLSDGVHVSATQAQNAVTTITDMLTHSTLU 1131  
 DB 1075 LIEARKYTGKILVYKQEWKGN--MDVSFKYCVSDGANTECSTCRFLHVLVEPEMLESFTL 1132  
 QY 1132 RLEDMSPERPLSLGLFLQAVATLAT--PRDHVVVFNQRPDTAPAGHILNVSLSVGP 1190  
 DB 1133 SLARMTVDQFMDPLVQFRDMASTLSWKSBDIHVIGKQHLD--DYIYINAIATDHGR 1190  
 QY 1191 PGPGGPPPLPSEDOQERLYLNRSLITAIISAQVLPFPDNLICLAREPCENYMCVSLRFD 1250  
 DB 1191 VVRGNRAIETLVESIKK---LEKMTLLQVEYIR-----DESGANFPCSMACRQTKQKV 1242  
 QY 1251 GSAPFIASSSVLPRIHPVGLRCPCPGFT-----GDYCEYVDLCISRPCGPHGRCS 1305  
 DB 1243 GEMKAHEIDNFAFRTLVNTVTVCECPGSGFTSSGAHGD-CDTRIDECYRGRSNNSTCYA 1301  
 QY 1306 REGGYTCLCRDGYTEHCEVSARSRCRPGVCKNGGTCCNLLVGPFC-DGP-SGDFEKP 1363  
 DB 1302 FENYTOCEKCPWMIGRHEISYHALTCVPGYMSISLC-ELDGNQMKCHCKYHEDIDE 1360  
 QY 1364 YQVYTRSPHASFITFGRLQRFAFTLASPAIKERDGLLLYNGRFNEKHDVFALEVIQ 1423  
 DB 1361 RCRIRSVSPDGGGLANVLDLPRQTQMTKFRSTIAHGVAVFTG--DRSRDFEVSVD 1418  
 QY 1424 EGVQLTFSNG--ESTTYSPPFPGVSDGQHTVQATKYNKLQDTGLPQSPSEQKAV 1481  
 DB 1419 RVIAVQFSLGGEKIDARKMENDVENRINDGEHWTALVYSNKQI-----T 1462



QY 1482 VYVGGCDYVALRFGSVLGNYSCAAQGNQGGGKKS-----LDLGTGELLGVPDL 1531  
 DB 1463 MSIDDCCTNSLILANT---SPNCAIRAKNALEKKCEPTYPCTRIYLDISNGLEFGGRPGT 1519  
 QY 1532 PESPPVPMRQFVGCGRNLQVDSRHIDMAF--IANNQTV-PGCPAKKNVCDN--TCANG 1586  
 DB 1520 SKQIE---KAFSGCISDLSDYDKEVDPESTIKEMHKVQVHEGCKHRDPCSTSDQCSAT 1576  
 QY 1587 GTGNQMDAPSCBCEPLDGFQKSCQAEANQHLGSSLVAMHGLS----PIQOPWLSL 1642  
 DB 1577 SKCNMRGRI CSCPCSVHSTGECVGLGTDLRGHSLFEBESFVLQPSQVSPFEVSF 1636  
 QY 1643 MPTROQADGVLQAITRGSTITLQREGHVMSLV-EGT---GLQASLRLBGRANDGD 1698  
 DB 1637 EFTSRAD---MQVF-----ALBPTORSVHYNLEVDQDTLKVINGBSEVELPAPETYSKI 1688  
 QY 1699 WHHAQALGASGGPRAHILSPDYGOQRAEENLGRHLGLSNITWGGITGPAGVARGF 1758  
 DB 1689 MMWVVIKFEAD-----SVATISINGIYSAEAKAS--ISDMNLESLYFGIAGT--GHPSRF 1739  
 QY 1759 RGCLQGRVSDTPEGVNSLDPSHGESINVEQ-----GCSLDPDPCD-SNRCPANYSND 1811  
 DB 1740 EGCIRNLVD-----GRSISVKKKGKTRAGCVVPMKCSVDISICPABS----- 1781  
 QY 1812 WDSVSCSDPGYVGDNCTNVCDL-NPCEHOSVCTRKSPADHGYTCECPRYVLGPYCETRI 1870  
 DB 1782 -----DITLPCVSANVNCSSGCVSNNTA--GYEICIPAKTKNQLEA 1825  
 QY 1871 -DQCPRGWNG-HPTGCPNCNDVSKGDPDCKNTSGECHKENHPRPSPTCLDCY 1928  
 DB 1826 PKQMCPSGWMGTPEPRCRRCSCAQTKDYEAQCDKKTGACQCKKSIFSTING--CVKCEC-G 1882  
 QY 1929 TGSLSRVCDEPDQCPKPGVIGRQCR---DNFPAEVTNNGEVVYDSCPRAIRGIV 1965  
 DB 1883 FQADSTECG-ADGCKCKNGAVGRCDRCRFBQDLKTLK-CRPVSGKCPSEIETSIQ 1940  
 QY 1986 WPTRFGLPAAAPCPKSGFETAHNCDEHKGWLPENLFNCTSIYFSELKFAERLQPNES 2045  
 DB 1941 WPASQKSIYRQSCPVGESGLATRKCLJTRGMSDVANMNCITRPREYSIMVAKFILE---- 1996  
 QY 2046 GLDSGSGQQLALLRNATQTAGYFGSDVAVYQALATRLAHSTQRGFSLATQVHYT 2105  
 DB 1997 -----PSKILITWANTNTSSIRGNQOIAEALSRLDVYRQSPMKRRAHIMKMKPT 2050  
 QY 2106 ENLLR-VGSALDLDTANKHWEILQCTEGGTAMLLQHTAEVAASLAQNMRTYVSPETIV 2164  
 DB 2051 EKLIESLGRVMSQEPADYSTILSK-----LNNYAEVVAETIHENV--NPLSPFVIAN 2100  
 QY 2165 PNIVISVRLDKGNFAGAKLPRYEALRGQRPDLLETVIIPESVRETPEVVRPAGGEA 2224  
 DB 2101 DHIYFADKLDPGNI---LPKFNNF-----VDLRPTGF-----PRVAVIYVGT 2141  
 QY 2225 QEBEELARORRHRELQGEAVASVITYRTLGLIPPNYDPDKSLAVPKRPILNPPVS 2284  
 DB 2142 Q-----VVSISIVY-----PRCNCRCENPMIA 2162  
 QY 2285 ISVADDEELLRALDKFVTYQFRLLETERTKPICVPMNHSILVSGNSARGACEVVR 2344  
 DB 2163 IVANTSD-----FVIVEFELBEEDGMKTEPCVAFDEK-----SGTWTAGALLIGL 2208  
 QY 2345 NESHSVQCNHMTSPFVADSVSRRENGEILPLKLTLYVAL-GVTLAALLTFPFLTLART 2403  
 DB 2209 NLTHACEYRIRIGFTWFDVQSSSIVRAVQMDMTSPALAGVALFLCFSLISLTLSSRS 2268  
 QY 2404 LRNSQIGIRNLTALALQAVPL--LGINQADLPFCYVITAILHLHYLICTSSWALTEA 2461  
 DB 2269 LKTHSVRIGLIFPALINILNLFVHKTAINDA---YCPRNAMLSSTSSAPAPMFLY 2324  
 QY 2462 LHLVRLTEVADVNTGMRFTYMLGWSVPATIGLAVG-----LDPEGSGNDPCWLSI 2515  
 DB 2325 LYIYRMLAD-----GSSSPGLTSLVGVIPPCILISTTTFVTVDCSLSP 2369

QY 2516 YDTLWSPAGPVAFAVMSVFLYIILARASCAAQGFKKKPGVSGIAPSVALLLSAT 2575  
 DB 2370 HLMFLWFCILLPGLPLLSFYAASVLSLKKVDVFAKKNVRAVQHLITFTIG 2429  
 QY 2576 WLLALSVNSDTLLPHYLPATNCIQGPFIPLSYVLSKVKALKACSRPS---PD 2631  
 DB 2430 MTLTGLFANQLPLPMEIMESIOSII---YLIALVFLWCVDITTKADSNPSMWLDNQ 2486  
 QY 2632 PALTSTLTSISYVNCSPYADR-----LVQPYDSAGLSHSTRSQSGOP----- 2677  
 DB 2487 KVMWESTVADP-QCASPLSPRHQHEVPMDSWVPVPSNHTH-TSINEBDTPRLL 2544  
 QY 2678 -----SYIPFLRBSALNPGQPPGADPGSL--FLBSQDQOHD 2715  
 DB 2545 LQGNRVINILSSPDQILNEGIVHYRNMGSLPLRBSAQDEAD 2589

RESULT 3  
 ID 08MY78 PRELIMINARY; PRT; 2909 AA.  
 AC 08MY78;  
 DT 01-OCT-2002 (TREMblrel. 22, Created)  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Ap-cadherin.  
 GN APCAD.  
 OS Aeterna pectinifera (Starfish).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;  
 OC Asteroidea; Valvatacea; Valvatida; Ascerinidae; Aeterna.  
 OC NCBI\_TaxId=7594;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Satch N., Humphreys T.,  
 RA Zhang S., Tsukita S.;  
 RT "A novel amphioxus cadherin that localizes to epithelial adherens  
 RT junctions has an unusual domain organization with implications for  
 RT chordate phylogeny.";  
 RL Evol. Dev. 0:0-0(2002).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREPREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC DB EMBL; AB075365; BAC06834.1; -;  
 DR InterPro: IPR000152; Asx hydroxyl.  
 DR InterPro: IPR002126; Cadherin.  
 DR InterPro: IPR000233; Cadherin\_C\_term.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR006210; IEGF.  
 DR InterPro: IPR001791; laminin\_G.  
 DR Pfam: PR00028; cadherin\_15.  
 DR Pfam: PF01049; Cadherin\_C\_term; 1.  
 DR Pfam: PF00008; EGF; 4.  
 DR Pfam: PF00054; laminin\_G; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 17.  
 DR SMART; SM00181; EGF; 4.  
 DR SMART; SM00179; EGF\_CA; 4.  
 DR SMART; SM00282; LamG; 2.  
 DR PROSITE; PS00010; ASX HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS00268; CADHERIN\_2; 17.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein.  
 SQ SEQUENCE 2909 AA; 316542 MW; 2466441BC3619CF CRC64;

Query Match 9.9%; Score 1538; DB 5; Length 2909;  
 Best Local Similarity 29.1%; Pred. No. 1.4e-87;

Matches 525; Conservative 235; Mismatches 792; Indels 254; Gaps 67;

QY 182 OPSVQATYPE-----NCPAGTVPASLRADIDEBGAGRLXTMDALFDSRNGPF 232  
 DB 1050 QPPEWSTPFDITWMENISINDVIGILOQOSRS-----GNGVYRVVQCOIQPTMSBGF 1104  
 QY 233 SL---DPVTGAVTTAEELDRETKSTHVRVTAQDHGMPRRSALATLTLLVTDYNDHDPV 288  
 DB 1105 ANQWVDSMIANLYISAPLDYETTKFTYLOVEAYDDG-NRLSLTIGQRVVWQDVNDDETQ 1163  
 QY 289 PQQOEKESLRLELVEGYVLTVRATDGPAPRN-ANILRLRLESGSGSSEVEFLDPBG 347  
 DB 1164 PVTNFEVAFPEVDVAPFLVTRVQADADTEAFKTYTSL--DQTFSAQYFRIDPPTGG 1221  
 QY 348 VTRTGPVPREVEEYQLTVEASDQGRDP-----GPRSTTAAFVLSVEDNDNNAPOPSBK 402  
 DB 1222 DLSTIKFPRSESDTSVQVVIATDGPSSLPSAGSNKGYLKVIVNVDKNDPFPDBP 1281  
 QY 403 RYVVOVREDVTPGAPVLARVTASDRDKSNVAVHYSIMSGNARQOFTLDAQTGALDVVSP 462  
 DB 1282 MYTRTIREDEBPVGYEVIGVATFVDPDS--IPRYLTANNNGAFEVDPATGALSIASPL 1339  
 QY 463 DVETTKETLRRAOQGRPELNSGLVTVQVLDINDNAPITVSTPQATYLESP-IG 521  
 DB 1340 DVEIOREYVWLYSANDG---LNVASTVRIQLENNDEKPEFELPVYTAEYSENDPNVP 1395  
 QY 522 YLVAVQALDADA-GDNARLEYRLAGVGHDPPTLNGTGTWISVAELDRBEVDYFSPGV 580  
 DB 1396 ROLLOVTAVDGDADVAQDIRSLEBGTGCTTITDAVQITLTETMLRESVSIVKLLV 1455  
 QY 581 EARDHG--TPALTASASVSVTVLDVNDNPTFOPEYTVRLMEDAVGTSVVTVSAND 638  
 DB 1456 KATDGGGTGTSITGVADLEVTVLDENDGPFPELETRGSPENSPGTFVWQVTAIDYD 1515  
 QY 639 ---AASVITYQ--ITSGNTN---RSTISQSGGGLVSLALPLDYLKLEQYLAATASG 690  
 DB 1516 DNGQGRITYSTISSEIPNDSSGLFQIDSGTGMVTVKSGAVLDRKNDYTLRLRASV 1575  
 QY 691 TRQDT-AQIVWVNTDANTHRPVFQSSHYTVNNEDEPAGTVVLLSATDETGEMARITY 749  
 DB 1576 PHEALTDIAIQLLDVNDNAPVFGGPVATSEBQPVATWERSYTDADVDFBEVOP 1635  
 QY 750 -FMEDSI-POFRIDAD---TGAVTTQAELEVD--QVSYTLAITARDNGIPKSDTYLE 802  
 DB 1636 GIIIGVGGSFELVADPLTLIGIRIARPLDYSPKLYTLITVDNNGSPGSGQTTT 1695  
 QY 803 ILYVNDVNDAPOQLDSTY-QGSVYEDVPFTSVLQISATDRDSCANGRFYTPGQDGD 861  
 DB 1696 IDVLVNDVAPTFNPALVYSAVNLBSVDVGTISIGTVSADPPEAGVFGQTFPSIDPASDP 1755  
 QY 862 GDFIVESTSGI-----VRTLRLDRENVAOYVLRVAVDQKMPARTPMEVTVTVLV 914  
 DB 1756 GEFSEIEISPIGAVNRADYKAKPLDRENVETHVLTLRATDVGPPPLGYATLANTLDDV 1815  
 QY 915 NDNPPVFEQDEPDVVEENSPIGLAVARVATDPDEGTNAQIMYQIVEGNIPVEFQDLIF 974  
 DB 1816 NDTPTPTFARD-YRPTIKEMLARVQFVYISAKQEDP--TGGPPTLVLENT-----DTSGF 1869  
 QY 975 S-----GELTA-----LVDDYEDRPREYVLTQATSAPLYSKATVH---VLLDRDN 1019  
 DB 1870 TWFVVGTTGNMNLISSKAISDPRETOATYSIPIRITDAPNGSGVNLVLYEIDBEND 1929  
 QY 1020 PVLGNFELFNPNVYTNRSSSPFGAIGRVPAHDPDISLTYSPERGNELSLV----- 1074  
 DB 1930 PHFGTTKELVYSFBN-----IPDSIGVGVDDKDTLEDKTY-----PELSDVYKQYR 1981  
 QY 1075 LNASGTELKLSRALDNNRPLKAIMSVLSD-GVHS-VTAQCALRVYTIITDEMLTHSTLR 1132  
 DB 1982 VVEDTGQVYI---LAGTPAGTYEFVDRVADGVHADQSTVIVDVADIPREAVRSSGSPR 2038  
 QY 1133 LEDMSEPERLS-----PLDGLFIOAVNATLATPPDHVVVNNQORDDAPGHLINSL 1185  
 DB 2039 FEGVTAEBELISVPSGGGQTKDLKGLIABEIIIPAKENVDIFSIVNVPQ--TTVDVRY 2096

QY 1186 SVQGPBGGGPPFLPSEDLQERLYLANSLLTAISAGVLPEDDNICLR-PCENYMCV 1244  
 DB 2097 S-----AHGSPYTPPQMGALANDRRIAQALGVITIGMIKIDMLLESACES--ACT 2147  
 QY 1245 SYLRPDSAPPFLASSSVLFRPIHPVGLRCRCPPGFTDYCETEVDLCYSRPGHGR 1304  
 DB 2148 NVLBIIDPQVYINTPSASFAVTVTSVAKVC----- 2179  
 QY 1305 SREGYTLGCRDGYTGHCERYSABSGRCTPVCNKGCTCNALLVGGFPCCDSSGPEKPY 1364  
 DB 2180 ---GANT-----QQPGLCDPACLNGGCTTETQGGGHTCCQCPG-YDGD 2220  
 QY 1365 COVTRSPFASPIFRGLRORPHFTLALSPATKERDGLLNGRPF-----NEKDFVAL 1419  
 DB 2221 CQQTTRF-XGFRHFGTLQOCBEHTSLBRTTTRABEVLLXNGMIVYTGMPDPMI 2279  
 QY 1420 EYIQEQVQLTFSAGESTTVSPFVPGVSDQOMHTVQLKYNKPRLLGQTLPOGSEQV 1479  
 DB 2280 QLLGGQPVLEINTAGSTLSLSPATNLGDCKMRLDY-YRN-----GKD 2323  
 QY 1480 AVTVYDGC-DTQVA-LRGSVLANVSCAAQSTGGSKSLDLTGPLLQGVPLDESP- 1536  
 DB 2324 VEFMLDRCKDAVALEASTSSRQTDQCKITQAPEDNKFVLVNTPLQIGI-DKSAGFTY 2382  
 QY 1537 ---VRMGQFVGCNRNLQVDSRHIDMADFLANNGTVPCCPAKQVCDSENTGNGGTCVNO 1592  
 DB 2383 PDGITPAASVYGCNRNLEQDSTLYDLTEPGAUVSBAQ--SKLYC--GECNN-GTCGCD 2437  
 QY 1593 WDAPSCBCEPLGFGKSCAQMANNPQHFAGSSLVAMHGLSLPI--SQPWYLSMERTQA 1649  
 DB 2438 FSTYCLCDPQYTBKCDQ--TPRAYDPATSSYVYQKPTVIDSRGNVQVSPRTED 2495  
 QY 1650 DGVLLQAT--RGSSTTLQRBGVMLSVGTLQASLLEBRANDGMHHAQLAG 1707  
 DB 2496 EGLMTWTNLNGLEFTTLKELINGYIR-SRWNLQDGEHSMYLDQYAVNDGMMHVNLE-- 2551  
 QY 1708 ASGGRHALISFD-YGQORAGNIGPR-----LHGHL--SNITVGGIIPGAPGVAR 1756  
 DB 2552 -----RFDTYITRIDDGGGVYRMESREBFTSLVEDPNSLLGAFVHVHVDISQ 2601  
 QY 1757 GFRGLQGVRSVDPREGVNSLDPSHGBS-----INVEQCSLPDPCDNPSPANSYC 1808  
 DB 2602 DFGGMDPRINNYLAFET-----ESDYAIPQPTASVTEGCSYDCAANPCVPFIC 2655  
 QY 1809 SMDMSYSCSDPRY--YGDCTNV--CDLAPCEHQSCTKPKPAPHGYTCBEPNYGP 1864  
 DB 2656 KDVWRKXICICKPGBELGNTCVAILDCCDPNCLNGICTDRDS--GFBCECPDGYRGD 2712  
 QY 1865 YCETRI 1870  
 DB 2713 ICDVAV 2718

RESULT 4  
 Q9WU10 PRELIMINARY; PRT; 4589 AA.  
 AC Q9WU10:  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
 DE Protocadherin.  
 GN FAT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=99173791; PubMed=10072790;  
 RA Ponsas M., Jacques T.S., Clanti U., French Constant C.  
 RT "Expression of the rat homologue of the Drosophila fat tumour

RT suppressor gene."  
 RL Mech. Dev. 80:207-212(1999).  
 DR EMBL; AF100960; AAD20459.1;  
 DR HSSP; P01132; 1EGF.  
 DR InterPro; IPR000152; Aex\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00028; cadherin; 33.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF00054; Laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 31.  
 DR SMART; SM00179; EGF\_Ca; 1.  
 DR SMART; SM00282; Lamm; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 17.  
 DR PROSITE; PS00268; CADHERIN\_2; 33.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_Ca; 1.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 1.  
 KW Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein.  
 SQ SEQUENCE 4589 AA; 505987 MW; 8CD0561BD58677 CRC64;

Query Match 9.8%; Score 1520; DB 11; Length 4589;  
 Best Local Similarity 26.6%; Pred. No. 4.5e-86;  
 Matches 478; Conservative 242; Mismatches 682; Indels 394; Gaps 57;

179 POFPPSYQATVPENQAGTPVVALRAIDEGSAGRLLEYTM--DALFDSRSNQPSLDP 236  
 2702 PRSEPPYTSISEDMPIGTEIDLIR-----EHSQVLTIVLKGNPESNRBEFVLDK 2756  
 237 VTGAVTAAEELDEETSTHYF---RVTAODHGMPPRSALATITLVTNDHPVPEQ 292  
 2757 QSRQLKLEKSLDHEHTKMYQFSILARCTLDYEV--VASIDVSIQVKANDNSPVLESN 2813  
 293 EYKESIRENLEVGVEVLTARATDGDAPNANITLYRLLEGSGSPSEVEFELDPPRSGLTR 352  
 2814 PYEAFIYENLPASRYVQVASHDSDGYNQVMTSLDOSODADIIISFALIMETGMITTL 2873  
 353 GPVDRREVESYQLTVEASDQGRDPGRSTTAAYFLSYEDNDNAPOSEKRYVVOYREYV 412  
 2874 KEIDHEERASYQIKVASDHG-EKVQLSSTALVDVTVTVDNSPRTAIIYGTISEDD 2932  
 413 TPQAPVLKATASDRDGS-NAVVHYISMSGNAGQFLDAQTG--ALDVVSPIDYETKE 469  
 2933 PPGGVALLSTTTADSEERINROVSYPITGDDALGQFAVENMOMDWRVYVKKPLDRREQDS 2992  
 470 YTLRVRAODGGRPLSVSGLVTVQVADINDNAPIFSTPFOATVLESVFLGYLVLYVQA 529  
 2993 YLLITVATDG---TSSKARAVEYKVDANDNSVPCRTIYSDAIFEDALPGCLVMQVSA 3048  
 530 IDADAGDAALELYRLAGVGHDPFTTINGTGWTISVAEELDRREYDFYSPVEARDHGTPA 589  
 3049 TDADIRENAETITVLPESGAR-KPKLNPDTGELRTLLALDRREGAVHLLVKAITDGG--G 3105  
 590 LITASASVTVLVNDNNPFTPTQPEYTVRLANEDAAVTSVTVVTSANDRDA--HSVITYQI 647  
 3106 RSCQATVTLTEVDNNDTPFTADPVAITVEFENTEPITPLRVOATDADAGLRKISYSL 3165  
 648 TSGNTRRFSTISOGGGLVSLALPLDYKLEQVFLAVTASD-GTRQD---TAQIVVNT 703  
 3166 LD-SADQFSLNRS--GIILQLEKHLDRLEIQAIVTTLTKAADQGLPRKLTATGIVVSVL 3222  
 704 DANTRHVPQSSHVTVVNNEDRPAGTTVVLISATDEYGENARITY--FMEDSIPOFRID 761  
 3223 DINDNPVFEYREYREGATVSEDIIVGTETVQVYASRDIRANAEITVAILISGNEHKGKSID 3282  
 762 ADTCAVTTQALEVDQVSTLAIATARDNGIPQKSDTITYLEILVNDVNDNAQPLFADSYQ 821

3283 SKTGAIPIIENLDYESSHGTYLTVEATDGTGTSLSDVATVNTINDINSVFGSDYST 3342  
 822 GSVEYEVPPFTSVLQISATDRDSGLANGRVFYTFQGGDDGDPFIYESTSGIYTRLRDLR 881  
 3343 TVVSEDALEQPVITIMDDADGPNSSHILYIISGNQ--SPFTIDPVAGEIKVKKPLDR 3401  
 882 ENVAQVLAIVAVDDKGMPPARTPMEVTVVLVNDNPPVFEDEDFVFEENSPIGLAVA 941  
 3402 EITISGTVTLVQAADNGNPPRVVTTVINDVSDVNDNAPLFSIDNYSVIIQENKPGFSTL 3461  
 942 RVTAIDPEGTNA-QIMQIVEGNIPVVFQDIFSGELTALVDLDEBPEYVLVIAQ-- 998  
 3462 KLVVTDKDSHNGPPPSFAIVSGNDNMEVQHGVLTA-ATVRKAVDHYLALVKVAD 3520  
 999 TSAPLVSRAT-VHVLDRNDPVPVLGNFELLFNNTVTRSSPFGAIGRAPADPDLS 1057  
 3521 NKPQSLSTHIDIRATESHPALPLEI---FITASGEYSGGVGKIHATDQDVY 3576  
 1058 DELTYSFERGNELSVILNASTGBELKSPALDNNPLEAIMSVLSGVASTAOCALRV 1117  
 3577 DLTYSIDPRND-GLPVSSSTGGKLIARHKDIG---QYLVAVSYTGKFTVADITVHI 3632  
 1118 TITTBMTLTHSITLIEDMSPERFLSPILGFIAVAATLATPPHVVENVQRTDADG 1177  
 3633 QQVTEMLNHTIATIRFANLTPEEFVGDVWRNFQALRNITGIKNDIQVSIQ--PSEBP 3690  
 1178 GHILVNSVSGPPGCGGPPFLBEDLOERYLNRSLTALSAQVLPFDNICLRBPC 1237  
 3691 SH-LDVLLEVK-----SGTQISTKQLHKNSSVTDBEILIGVRLIDFQKLCGLDC 3744  
 1238 -----ENYRCVSVLRPDSAPFIASSSVLPFRPHVGLRCR---CPGPGFG 1282  
 3745 PKFCDEKVSVDENIMSTHSTARLSFTVRHRTAVCL-----CKDGTCPVHVG 3794  
 1283 DYCEBTVDLCTSRPGCPHRCSS--REGYTTLCEPDGTIGECVVSASGRCTPGVCNKG 1340  
 3795 -----CEDNPPAGSECAVDAREERYCVCV-----PG----- 3821  
 1341 GTCVNLVVGFCDDCPSGDFEKPQCVQVTTSPSPASFTFRGL--RQRFHTALSFAIK 1398  
 3822 -----GGF-AKPGS-----SITTTGSSPVYKRLMBENRLEMLTMRIRY 3863  
 1399 EEDGILLYNGRNEKHFVVALEIOVQLTFESAGESTTVTSPFPVPGVS-----DQO 1451  
 3864 SSHAVVMY---ARGDYSLIEIHTRQLYKEDCSG-----PGLVSQSIQVNDQO 3911  
 1452 WHTVQLKYNKRELQGTGLPOPSQKVAVTVVDGCDTGVALRFGSVLGNVCAAGTQO 1511  
 3912 WHAVSLV-----EGNYAKVLDEVHT-----ASGTAP 3939  
 1512 GSKKSLDLTGPLLGGVPDLPESPFVRMQ-----FVGCWRNLQVDSRIAD 1557  
 3940 GALKTLTLDNVTYIFGG-----HLRQGTGKGRNTQVANGRCG-----DSIYLN 3984  
 1558 MADFIANN-----GTVPCC--PAKNVCDSDNTGANGTC--VNQMDAFSCBCEPL 1602  
 3985 GQELPLNNPRAVAHLEBVVDLAHGLLTATBD--CSSNFCQNGGVCPNPTGYCKCSA 4043  
 1603 GFGKSCCAQEMANPQHFLGASSLVANHGSLPISQPMYLSIMRTQADGVLLQAITRGSS 1662  
 4044 LKAGTYC----- 4050  
 1663 TITLQLEBGMVLSYEGTGLQASSIRLEGRANDGDMMHQAQLAGSGPGHAILSPDYG 1722  
 4051 -----EVSVPNCSSN----- 4060  
 1723 QQPARBGNLPRHLGLLSNITVYGIPGAGYARAGRGCLQGVAVSOTPEGVNSLDPSHG 1782  
 4061 -----PCLYG-----GTGVNDNGGVCCQCRGLYTG----- 4085  
 1783 ESINVEGGSLLPDPDSDNCPANSYCSNDMDSVSGSCDPGGYVGDNC---TNYCDLNPCEH 1839

Db 4086 -----ORCOLSPYCDPCKNGTCTPDSLDGAVCCDSGFRGRCOSIDDECAGNECRN 4139

QY 1840 QSVCTKRPSPAHG-YTCECPNLYGPYCETRIDOPCPRGWGHPTGCPNCNCKVSKG 1894

Db 4140 GALTEN-----THSGHCNCSOEYRGKHCEDATP-----NHVSPFNMIGLALG 4183

RESULT 5

Q8ABX7 PRELIMINARY; PRT; 4610 AA.

AC Q8ABX7; 01-MAR-2003 (Tremblrel, 23, Created)

DT 01-MAR-2003 (Tremblrel, 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel, 23, Last annotation update)

DE Protocadherin.

GN FAT.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

RN (1)

RP SEQUENCE FROM N.A.

RA Down M.L., Ralston K., Smith S.I., Burns G.F., Boyd A.W., Power M.;

RT "Zebrafish Fat Protocadherin."

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF469608; AAO15696.1; -

SQ SEQUENCE 4610 AA; 506628 MW; 699B20BA4FD63B93 CRC64;

Query Match 9.8%; Score 1519; DB 13; Length 4610;

Best Local Similarity 24.2%; Pred. No. 5.2e-86;

Matches 500; Conservative 248; Mismatches 674; Indels 640; Gaps 52;

QY 174 NNTAPOPPOPSYQATVPENOPAGTPVASLRATDPGEAGRLTETDALPDRSNQPS 233

Db 2381 NDVT-PVFSQNYEATVSELRPHGFYTOVQASDADISGKLF---SILSGNEQNRA 2436

QY 234 LDPVTAQVTAELDETKSTHYFRYTAODHMPRSALATLTLLVTDTNDHDPVFEQ 293

Db 2437 MDNTALVTSNRKPMESLIYNINVSVD-GVFRSAIYKAVNI--SANFHPPTNQVD 2493

QY 294 YKESLRENLVGEVLTVRATDGDAPNNAILYRLLEGSGSPSEPEIDPRSGVIRTG 353

Db 2494 YVVELLENSPVGLVLAQAQITDDSGYKLIHYINDIAKDFSISE---NGEITFLE 2549

QY 354 PVDRBEVESYQLTVEASDGRDPSPSTTAFLSVEDDNDNAPQFSKRYVQVEDVT 413

Db 2550 SLDR--NALEKVIPISLAKOGGKVGFCYVIVILTDINDNAPQFRADYRVNVAVDV 2607

QY 414 PGPVLRVTRASDDKGSNAVHYSI-----NSGN----- 442

Db 2608 RGTTLIKIASDMDGSSNADITYSIETDADNVEGNPEIHQPSGVIYKESLIGLENQLYT 2667

QY 443 ----- 442

Db 2668 FLVRADGGNPTSSVVPVYVKILAEVVPKFIESHYRAIEDSLSGEIDVITGTS 2727

QY 443 -----ARGQ-----FYLDAQTGALDVVSPDYETTKETTLRVRAQD---GGRPP 484

Db 2728 QPVFSLVKGNTESNODBEVVDPSGSLKDKDHBSTKMYQTLTLDQSSYBSKV 2787

QY 485 SNVSGLVTVQVLDINNAPIFVSTPQATYLVESVPVGLVLAQALDADAGNARLEYRL 544

Db 2788 SAAD--ISIQKQVNDNRPLFESNPYEAVENTLPGTCVIOVKATDLDSTNGHYVNL 2845

QY 545 AGVGHDP-----FTINNGTGWISVAELDEBEVDVFSFGVEARDHGTFA--LTAS 593

Db 2846 D-----PNEGTRDIYELFAINSESGMITTLKELDETRKTYVSLADRGKQULMS 2899

QY 594 ASVSVTLVDVNDNPPFTQPEYTVRLNEDAAVGTSVVTVSAVDRDA---HSVITYQITSG 650

Db 2900 TRVDVTVVDVNDNPPKFTAEIKGTVSBDDPGVIALISTTDLTEENNKQISTYITGG 2959

QY 651 NTRNRPSTISQSGGLVSLALPLDYTLERQVLAVTASDGTRODTAQIVNVNTDANTHR 710

Db 2960 DPLGQFALEHTQEMKVSVRKPLDRREKONYLINTATGTGFAKAVEKVLADANS 3019

QY 711 VPQSSHYTVANVEDRAGTTVVLLSATDEDTGENNATY--FNEDSLPQFRIDDTGA 769

Db 3020 ICKESFYMESVPEDSPAGRLILQVSAITDADISNMAISYELLDPGSEHSISDETEBLK 3079

QY 770 QALDEYDQSVYLAITATDNGIPOKSDPTYLEILVNDNDNAPQFLRDSYQSVYEDV 829

Db 3080 LPLDREBEERVKMKRALDGG--GRPCAEVETIYEDVNDNPPQTTPTTYTVEENTE 3137

QY 830 PFTSVLQISATDRDSGLNGRVFTYFGDGDGDFIVESTSGIVRTLRLRENNVAQVYL 889

Db 3138 ITPVRLYASDLDGTANARIMYSIL--DSADGVFSIEETGVVRLDRPLDRLOSLYTL 3195

QY 890 RAVAVDKMPARTPE--VTVVTVLDVNDNPPPEQDEFPVPEENSPIGLAARVATDP 948

Db 3196 RQATDRGSPRLHSQTLTSLVSLDINDNPPVERERRYATAEDIPVGTQVLRVHAAR 3255

QY 949 DEGTNAQIMQYVEGNIPEVPOLDIPSGELTALVDLDEDEPEYLVIOAT---SAPLVS 1005

Db 3256 DTEAGTEITVAILINGNERAFVVDPOTVGVPIETAEHFIETVETAGTGSLSLD 3315

QY 1006 RATHV----- 1011

Db 3316 LATVINILTDVNDSPVFNQDIYSAVISEDAELGKTVLTMADADGPFSSNQVRFIIG 3375

QY 1012 -----RLDR-----NDP 1020

Db 3376 NQSPFTIDPVGEVKAALLDRKTSYTLTVLASDNGPARSSATTINVDSDINDP 3435

QY 1021 PVLG-----NFEILFNN----- 1032

Db 3436 PIFSQANYSIIIOENQPIGASVQLTVSDRDSHNGPPTFTSIRGNBEDFRITPGVL 3495

QY 1033 -----YVTRSS 1039

Db 3496 VSAATLSRQTOEYLLQAVQVDSGRPLVSTAFVSRVLEESIYPPSILPLDIFITATD 3555

QY 1040 SFPGALGRVPAHDDISDLSYSP-----EKGNELSVLNLASGCELSLALDNNRL 1094

Db 3556 EYSGVGLKIAHTDQVDTLTYSILAPDSSSTSENSGLFSVPADOKLVARGNLDG--- 3612

QY 1095 EAIMSVLSDVSHVSTACALAVTITITDEMLTHSITLRLSDSPERFLSPILGLFIOAVA 1154

Db 3613 QVVLNITVTDGFTAAARNINHQATQALDNLSAVRSALAPBEFIDYVRNPLRALR 3672

QY 1155 ATLATPPDHVVFNQRTDAPQGHILNVLSVQPPCGGPPPLPSEDLOERYLNS 1214

Db 3673 NIAGVRGQVQLVSLQPADASGD--LEVLTALER-----SGSPQPOEVLRYKLNASVG 3725

QY 1215 LTLTAISQGVLPFPDNICLRBPCENYMGCVSLARDSSAPFLASSVLRPPIHPGLRC 1274

Db 3726 VLEEMTGRIYAVVKKLCAGLDLP--LSFQETIATETA--VASYTARISFTPRMRTA 3783

QY 1275 RCPPEFTDYTELEVLDVLSRCPGPHGR--RSRSGYVLCRDGYTGEHCVSARSGR 1332

Db 3784 KCP--CBAGBSILSKLCEGSPCPGMECVEDPTHSKRCLPDE---GQDECSBSRQSLT 3638

QY 1333 TRGVCKNGGTCVNLVGGFKCDPBGDFEKYPCQVTTSSFPNASHFTTFRGLRQHFYTLA 1392

Db 3839 FSG--NGARVRLMENERK-----EKMLSLRLRIFSTHAIVHQRLDH--- 3883

QY 1393 LSFATKERDGLLVNGRFRNEKADPVALFVIGQVQLTFSAGESTTVSPFVGVSQGM 1452

Db 3884 -----TEDNDGPIQYK-----FDGSGPGVSVSH--STVSDGGM 3916

QY 1453 HTVQLKYNNKPLLGQTLPGGSEQKVAVTVVGDGDTGVALRFGSVLNTS-----CA 1505

Db 3917 HTVSL-----VNGVYALVLDQVYA 3937

QY 1506 AQTGGGSKSLDNLGPIILGG--VPDLPSFPVRKQPVGCMRNQVDSRHLDMADFLAN 1564

DB 3938 ASGTAGTTLRTLNTDTSMTFGGHNVRPASGSGRLVMSGLRCLGIVNGHELP.LSGVRA 3997  
QY 1565 NGTV-----PSCPAKKNV--CDSNTCHNGTGC-VNOMDAFSCBCEPLGFGKSCAQEMA 1614  
DB 3998 HATLEELVAVAPCSGLAPRPGCTSNPCCTNGCTCSALPBGNGYCKCTAAFMGTHC----- 4052  
QY 1615 NPQHPLGSSLVAMHGLSLPISQPMYLSLMFRTRQADGVLLQALTRGRSTITTLQLRBGMV 1674  
DB 4053 ----- 4052  
QY 1675 LSVETGGLQASSLRLEPRANDGDWHHAQLALASGSGPHALISPDYGOARAGNIPRL 1734  
DB 4053 -----EVTISPCASN-----PCL 4065  
QY 1735 HGHLSNITVGGIIPGAGVARGF-----RGCLQGVVSDTPREGNSLDPHGESINVEOG 1790  
DB 4066 Y-----GGTCIRGG--DFYCCQCGQYSG-----OR 4089  
QY 1791 CSLPDPQSNPCPANSYCSNDMSYSQSCDPGYGDNCC--TNVCDLNPCEHQSVCTRRP 1847  
DB 4090 COLGPFCTNPPCKNSGKCIDSLDGPVCEBPGFQGBRCLSDVBCLNKPCNGHCONTRY 4149  
QY 1848 SAPHGTCGCPHYLGPYCETR 1869  
DB 4150 GS---PSCNCSNGYSGQHCELR 4168  
RESULT 6  
061230 PRELIMINARY; PRT; 2809 AA.  
AC 061230;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE G-cadherin.  
OC Lytechinus variegatus (Sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Temnopneustidae;  
OC Lytechinus.  
OX NCBI\_Taxid=7654;  
OX RN  
RP SOURCE FROM N.A.  
RX MEDLINE=98104238; PubMed=9441671;  
RA Miller J.R., McClellan D.R.;  
RT "Characterization of the role of cadherin in regulating cell adhesion during sea urchin development.";  
RT Dev. Biol. 192:323-339 (1997).  
RL -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DB HSPB; U54823; AAC06341.1; -  
DR HSPB; P01133; IEGP  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR002126; Cadherin.  
DR InterPro: IPR000233; Cadherin\_C\_term.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR006209; EGF\_like.  
DR InterPro: IPR006210; IEGF.  
DR InterPro: IPR001791; Laminin\_G.  
DR Pfam: PF00028; cadherin\_16.  
DR Pfam: PF01049; Cadherin\_C\_term; 1.  
DR Pfam: PF00008; EGF; 4.  
DR Pfam: PF00054; laminin\_G; 2.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 17.  
DR SMART; SM00181; EGF; 3.  
DR SMART; SM00282; LamG; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00232; CADHERIN\_1; 2.  
DR PROSITE; PS50266; CADHERIN\_2; 14.

DR PROSITE; PS00022; EGF\_1; 3.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS50025; Lam\_G\_DOMAIN; 2.  
KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
KW Glycoprotein  
SQ SEQUENCE 2809 AA; 303041 MW; 60191A3B643BA0F3 CRC64;  
Query Match 9.7%; Score 1508.5; DB 5; Length 2809;  
Best Local Similarity 29.0%; Pred. No. 1e-85;  
Matches 538; Conservative 252; Mismatches 757; Indels 311; Gaps 87;  
QY 147 SCKLAQAGLRAGE-----RSPBSLGGRRKRVNTAPQGPSPYQATVPEHNPAGRP 199  
DB 938 SCIVTASGIRSRVDVDFVRSPPDT-----TYGVPTFSNTPYQVTTINENTTIG-D 988  
QY 200 VASLRALIPDEGEARLEYTMDALPDSR-----SNOFSL-----DPVGAVTAAELDET 251  
DB 988 VLLVSV-----GAQGVGPDIVGSGSPSTNSDGTFTLTRDDMLQSDTLIVELDREA 1042  
QY 252 KSTHVFRTYADHGMPPRSALAT-LTILVTDTNDHPVPEQGEYKSLREMLEGYE--V 308  
DB 1043 VQEVELVTAQYVGVGQEFIAQTPTVTIEDFNVDVAPKFPATVFGYIAENQPVANSDDPI 1102  
QY 309 LTVRATDDDPAPNAILIRLLEGSGSPSEVEIDPRSGVTRTRGPVDEEVE--SYQLT 366  
DB 1103 ITLQATDDVTPPE---YSQISYSFVNABSPESIDEDGELFALVFPREALIEDSIALT 1158  
QY 367 VEASDQGRDPGPRSTTAAVFLSVEDDNDNAPQFSEKRYVQVREBVTGAPVLRVTASDR 426  
DB 1159 VRASD-----GVNQLATVFTIYDENDBEPFN-GTFSFVLELVGNGYDGTGTATDD 1212  
QY 427 DGKSNVAVHYSIMSGNARQFYLDAQALDVSPLEDTYETKEYTLRVRAQDGRPPLSN 486  
DB 1213 D--ISEVLEYFISGNEGGAFTVDAEGRIRAGVLDYEARTSYELQSVNDG-----KN 1265  
QY 487 V-SGLVTVQVLDINNAEIPVSTPRQATYLE--SVPLGVLYHQAQI--DADAGNARLE 541  
DB 1266 VATTVTNTVNVANVAAQPDQSAVSASVIEBDSNLPRILISVAATQGDADAVDA--V 1324  
QY 542 YRLAGVGHDFPPTINNGTGMISVAELDREEDVFSFGVEARDHGTPTALTSASVTVL 601  
DB 1325 YGLVGTGAGTIFTIDSGNITLTQLADRBHPIPTNLAAVANDNGNGLTSVVDVTEIE 1384  
QY 602 DVNDNPPFTQDEYTVRLNEDAVGTSVYTVASVDRDAHSVITYQITGNTNRPSITSQ 661  
DB 1385 DINDNAPYPPDQEVYGVSEKRPPTPVAVAAEPPDADLMYFPPSP--DNNINSQ 1442  
QY 662 SGGGLVSLALPVDYGL-BROYVLAVTASDGTRODPAQIVNNTDANTHRPVQSSHY-TV 719  
DB 1443 T--GQITTAQDFRRTPPSEYEIEVOATDGVNTASTVATISIDVDNKPSPSEBVTYPPDA 1500  
QY 720 NVNERDPAGTTVLLISATDEDTGENARITYFMEDSIPQ--FRIDADT-----GAVTTOAEL 773  
DB 1501 SVLETPETGTTITTVQALIDPDVDFRDAVFSINSDDPDELFRIVADAATLQGLIQVNEL 1560  
QY 774 DYEDQVS--YTLATARDNGIPQKS--DTTYLEILVNDVNDNAQPLRDSYQSVYEDV 828  
DB 1561 DETLATNEFTLTVAVTOSGPTDSGRPEIALVKILVKNANDLAVFQDPVDVAGVSBPA 1620  
QY 828 PFTSV--LQISATPRD---SGLNGRVFTPGGDDGDDGFVBSGTVRLRLADREN 883  
DB 1621 TNSQVGFPTSATIDGDEFGATLN---YIIDPTTDPQGFAINENQOLI-VASPLDRET 1675  
QY 884 VAQVYLRAYAVDKGMPARTMEVTVLVLDVNDNPVEEQEDFVFEENSPI--GLAVA 941  
DB 1676 VASYELKYYAVDNGSPMSGTVAVTVDVDTDPHQAQD--YNSVSEB-GRIBANVEV 1733  
QY 942 KYTATDPPE-GTNAQIMQIVGAINPEVFQDLF-----SGEL---TALVDLYE 987  
DB 1734 SYEAVDDDDPPSGPPLYNVA---PQPDWTTFPDIBGLGTSGSIRVSTTGLEIRE 1789  
QY 988 DRP--EYVLVIQTSAP--LVSRATVHVELLDNRNPNPVLGNFELLNNVYVNRSSSPFG 1043

```

Db 1790 THEPYDIVFLIAVETPEALTGOTITLMISDVNDNPHVATKOLLVSYEGN----1PT 1845
Qy 1044 GAIKRVPAHDPDISLTYTSFERGNEI-SLVILNASTGELKSRALDNRRPLEAI-MSVL 1101
Db 1846 TEVGKKGVEDPDLEKTY--EAVGILPDFQOLDSTGDTMA---EGTPAGVEMDR 1899
Qy 1102 VSDG--VHSTACALRVITITDEMTHSTTLLEDMSPERFLS-----1LAGFL 1149
Db 1900 VSDGTYESVSVIYTVKDIPREAVSSGSIRFSGTTLABELTPSEGSNDRLKAVIL 1959
Qy 1150 IQAVATLAPPHVYVNV---QRTDAP-GGHILNLSVQCPGPGGPPPLSEDL 1205
Db 1960 ANAVGQQLAN-PFIFSVLVNNGERTVDIRYAAH-----GSYYYPADL 2002
Qy 1206 QERLYNRSILTAISAOR-----VLFPDNICLAE-PCENYMRCSVYRPPSSAP 1254
Db 2003 D-----LAALSTSNELBELGLSLAQIPVD-LCVENNVES--SCNVLVVDPTPT 2049
Qy 1255 FLASSSVLFRPIHPVGLRCRCPPTGTYCETEVLDYSRPCGPHGRCSRREGYTCLC 1314
Db 2050 VVDSGTASLVAITSV--LEAQ-----ICGARVAF----- 2078
Qy 1315 ROGYGHEGEVNSRSCRTPGVKNGGTCVNLVGFKDCSGDPKRYCQVTTSPRA 1374
Db 2079 -----GNCSDPCLNGCTCEDVHGTYRCTCFY-LFDGPNCOQTKRSF-B 2121
Qy 1375 HSFTFRGRORPHFTALSPATKRDGLLVNGRF-----NEKHDFVALVTOGVOULT 1429
Db 2122 NGVASTIRROCEBLSLEIFTEVSSGILLNIGPFTPTGDDPIDMILELGGARLT 2181
Qy 1430 FSAGESSTT-----VSPVPGVSDGOWHTVOLKTYNRPGLQOTGLPOGPSOKVAVT 1483
Db 2182 INIGSTSDTDMDVLEAPTDPTDQINDNEMHRIDV-YRN-----GRVEMT 2225
Qy 1484 VDCGTGV---ALRFSSVLGNISCAACGTQSGSK-KSLDGLGPELLGV-PDLPSFPV 1537
Db 2226 VDRC-MGVFPFAETSSSTLDTSSCRVNGTTPGNNMLRFLNVHTPLFGMSADYDVTVP- 2283
Qy 1538 RMRQFVGCNRNLQVDSRHIDMADFIANNGTVPSCPAKKNVC-DSN---TCHNGTCVNM 1593
Db 2284 ---SGPDGCIKNLVSDGFLYDLGTPGTSSEBACPTDQCTDDNMPVCNN-GTCEADI 2340
Qy 1594 DAPSCBCLPAGFGKSCAQEMANPQHFGLSGLVAMHGLSLPI---SQPWLSLMFTROAD 1650
Db 2341 BSFICICPFGFNGLTCDVEL-TPYDPAISYITVELDSSLVDADRASSNYQLMVRTROEN 2399
Qy 1651 GULLQAITGRS--TITLQRBGHVLSVEGTGLQASSLRLRPGRANDGMH-----H 1701
Db 2400 G-LIMSISANTYETIRMEMVOGELKADWH-LGDKFVSVTMVFNSINDGAMHAINFRDXD 2457
Qy 1702 AQLALGASGPGHALISPDYGOORABGNLGRHLGHL-SNITVGGIPGPAGVARGPRG 1760
Db 2458 SVYTIKIDGGGKELQ-----NRSQYSGLDVENSILVGAFTVNTVTDDEFG 2507
Qy 1761 CLQGVAVSDT---PEGVNSLDSHGBSINVEQCSLPDPCDSNPSPAN-SYCSNMDSYS 1816
Db 2508 CANDRIRINFLMGEGINDYAVA-TKSAGVTEGCP-SDVCDSDPGCSILVCTDWMRFYE 2565
Qy 1817 GSCDDGY-----GNCNTVCUL--NPGHOSVCTRKBSAPHGYTCCECPNTLGYCE 1867
Db 2566 CLCPGQEEVEDDPTCAIIDCVNFCANGGTCVE--GDPGYTCDPCSGYGYDRCE 2621

```

RESULT 7  
090X3

ID 090X3 PRELIMINARY; PRT; 4587 AA.  
AC 090X3;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 13, Last annotation update)  
DE Mouse fat 1 cadherin (Fragment).  
GN PATH OR MPAT1.  
OS Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=embryo;
RA MEDLINE=20203461; PubMed=10741417;
RA Cox B.T.M., Hadjantonakis A.K., Collins J., Magee A.I.,
RT "Cloning and expression throughout mouse development of mfact a
RT homologue of the Drosophila tumour suppressor gene fat.";
RL Dev. Dyn. 217:233-240(2000).
DR EMBL; AJ250768; CAB65271.1; -.
DR HSSP; P08709; 1BP9.
DR MGD; MGI:109168; Fatch.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00028; cadherin; 33.
DR Pfam; PF00028; EGF; 4.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 31.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00282; LamG; 1.
DR PROSITE; PS00232; CADHERIN_1; 17.
DR PROSITE; PS00268; CADHERIN_2; 33.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS00025; LAM_G_DOMAIN; 1.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
FT NON TER 4587 4587
SQ SEQUENCE 4587 AA; 506036 MW; 4D3F23B05127CB4 CRC64;

```

Query Match 9.6%; Score 1496; DB 11; Length 4587;  
Best Local Similarity 23.5%; Pred. No. 1.5e-84;  
Matches 533; Conservative 267; Mismatches 662; Indels 810; Gaps 68;

```

Qy 179 POPOPSTQATVPENQAPAGTVASIRAIIDPEGEAGRLLEYTMDLAFDSRSNOFSLDVT 238
Db 2283 PVFVQPSISTTLSEASVGTIVLOVRATDSSEPNRGISYOLIG-NHSSKHDFHIDSENT 2341
Qy 239 GAVTTAEELDRKTSSTHFRVTAQDHGMPRASALATLITLVTDNDHPVVEQOEYKESL 298
Db 2342 GLISLVRLADYEQGQRIIFRAVDGGMPLSDSVVTVANTDLDNNEPLLEQGVYEAR 2401
Qy 299 RKNLEVGTEVLTVAATDGDAPPNANILYRLKSGSGSBESEVEFETDPPSGVT----- 349
Db 2402 SEHAHGHFWACVRACADSDSLDKLEYSIL---SGMDHKSFIIDRETGITLSNLRHT 2458
Qy 350 -----RTRGPV-----DREVE-----SYOLTVESAD 371
Db 2459 LKPYISLVNSVSDGVFRSSARVNVTVMGNLHSPVPHONEYEVELAENAPLHTLVQVKA 2518
Qy 372 QGRDPG-----PRSTAVFLSYE----- 390
Db 2519 SDRDSGIYSHVTHIIVNDPAKDRFYVNDROGIFTLKLDRETPTAKYISIRLMAKDAAGK 2578
Qy 391 -----DDNDNAPQSEKRYVQVREDVTGAPVLTAVTASDDKSGSNVAVHYSI- 438
Db 2579 VAFCTVAVIITDDNDNAPQFASSTKYEVNIGSSAAKGTSVKVFASDADBSNADVTYAI 2638
Qy 439 ----- 438
Db 2639 ADSESVKENLEINQLTGLITTKSLIGLENFFTFPRAVDSGSPRESVVPYIKILPP 2658
Qy 439 -----MSGNA-----RGQFY-LDAQ 452
Db 2699 EVQLPREFSEPPYTYTISEDPTIGTEIDILRVEHGAVLITLVKGNTPESNRDEFVDRQ 2758
Qy 453 TGLADVSPIDYETTKETYLKRVRAQDGRPLSN---VSGI-YTVQTLNDNAPIVST 508
Db 2759 NGRKLKESLDHETTKYQPSILA---RCLTDYEVVASIDVSIQVKDANDNSPVLESS 2814

```

QY 509 PFOATVLESVPLGVLVLAHVAIDADAGNARLERL-----AGVGHDPFTTNGTGMIS 563  
 Db 2815 PYEAFIVENLPQSGRVQIRASDLDSGANGVWYSLDQSOADIIIES--FAINMEKWT 2872  
 QY 564 VAALBREEVDYSPFGEARDHGTPA-LTASASVSVLVLDVNNNPTTQPEYVLANED 622  
 Db 2873 TKELBHEERASIOIKVVASDHGKVKQSLAVGVTVTDVNSPPFTMEIKGTVESD 2932  
 QY 623 AAVGTSVVTSVAVDRDASHV---ITYQITSGNTRNRPSTISOSGGGLVSLALPLDYKLR 679  
 Db 2933 DPGGVTAIISTDADTEINRQVSYFTGGDALGQAVENVOGDMRVYKFKLDRBQD 2992  
 QY 680 QYVLAVASDGTRODTAQIVNVTANTHRPFGSSHITVNNVEDRAGTTVVLISATDE 739  
 Db 2993 SYLLTVATDGTSSKRAVEVKVLADANDSPVCEKTSYSTDITPEDALPGELVWQVSATDA 3052  
 QY 740 DTGENARITYFMEBS-LPFRIDADGAVTTOALEYDEDOVSYLAITARDNGIPKSDT 798  
 Db 3053 DISNNAITTLTPOSGAKETKLPDTEIETLALDBEQAIVNLVKAIDGG--GRSCQ 3110  
 QY 799 TYLEILVNDVNDNAPOFLRDSYOGSVYEDVPFTSVLQISATDRSGLANGRVFTYRQGD 858  
 Db 3111 AALVLTLEDVNDNAPEFTAPRYTITVEPTEPGLTRVQATDADGLANKISYSLV--E 3168  
 QY 859 DGGDPFVVESTSGIVRLRLDRBNVAQYVLRAYNDKMPARTPM-EVTVVLDVNDN 917  
 Db 3169 SAQGQFISNRSGLIOLEKHLDBELQAVYTLTKAVDQGLPRRLTATGVVAVSLINDN 3228  
 QY 918 PVPFODEPDPVVEENSPGILAVARVATDPDEGTNAQVQVIEGNEIPEVFDLPSGE 977  
 Db 3229 PPFETREYKASVSEDVIGTEVLQVYASRDIENAKETIYALISGHEKFSIDSKTGA 3288  
 QY 978 LVALVDLYEDREPEYVLQAT---SAPLVSRATVHRLDRDNPVLG----- 1024  
 Db 3289 IFTIESLDYESSHYULTVEATDGTPLSDVATVINNTDINDSPVFSQDTYTTVSE 3348  
 QY 1025 ----- 1024  
 Db 3349 DAALBOVITIMADADAGPSNSHIYSILEGNOGSPFTIDPVRGEVKVTKPLRETTISG 3408  
 QY 1025 -----NPEI-----LP--NNY-----VTN 1036  
 Db 3409 TLTVQADNNGNPRVNTTNNIDVSDVNDNAPLFSRNTYSVIIQENKPVGSVLKLVVD 3468  
 QY 1037 RSSS----- 1040  
 Db 3469 KDSHNGRPFFFTIVSGNDENAEVNOHGVLTATIKRKYKHFFLVAVKADSGKQLS 3528  
 QY 1041 -----PPGALGVNPAHNDPDISLSTYSFERGN 1068  
 Db 3529 SMTHIDIRVIESIHPPAILPLEIETAFGEBSYSGVIGKIHTDDVYDTLMYSLDPHM 3588  
 QY 1069 ELISVLINASTGELKLSRALDNNRPLEAIVSVDSVHSVTACALRVITITDEMILHS 1128  
 Db 3589 D-GLFVSYSYSGKLIARHKDIDG---QYLLNVSTYDCKFTVADIVTHIQOVQEMLNT 3644  
 QY 1129 ITTLELMSPEPRLISPLGLPIQAVATLATPPDHVVVAVVNOVDTDAPGSHIIVSLVG 1188  
 Db 3645 VAIRFAMLTPEEFVGDYWRNFORALRNILGVKKNDIOIVSLQ--PSEFHSN-LDVLLFVE 3701  
 QY 1189 QPPEGGGPPPLBSEDOERLYNRSU--LTATSAQVLFPPDNICLRBSC----- 1237  
 Db 3702 R-----SGTTFTSKQLHK--INSSYTDVEIIGVILLEVQKLGGLDPCPMKFCDEKV 3754  
 QY 1238 -----ENYVRCSVLAIPDSAPFIASSSVLFRPIHVGIGRCR---CPRGFTGDYCETEVDL 1291  
 Db 3755 SVDENVMVSTHSTALRSFVTPRHRTAVCL-----CKDGCPRVHNG----- 3795  
 QY 1292 CYANPCGPHRCMS--REGGYTCLCRDGYGHEHCVSABSGRCCTPGVCNKGICVALLVG 1349  
 Db 3796 CEDNPCGASBCVADPREEKYSVCV-----PG-----G 3823  
 QY 1350 GFKCDSPSGDPBEKVCVTTTRSPFASHTFRGL--RQRFHPTLASFATKERDGLLLY- 1406

Db 3824 GFG-KCPGS-----SSTFGNSFYKYLLINENRLEMKLSMRLATYSIAHAVMYA 3873  
 QY 1407 NGRFNKBDPVALVLEIYQEVQVLTFSAGSTTVSPFVPGVSDGQMTYOLKYKPLLG 1466  
 Db 3874 KGTDYISIPGISVQIQ-----VNDGQMHVASEV----- 3903  
 QY 1467 QTLPGQSEBQKVAVTVVDCPTGVALRFGSVLGNYSCAAQGTQGSKSLDLTPGLLLG 1526  
 Db 3904 -----EGNYALVLVDEVHT-----ASGTALGALKTLNLDNVVFG 3938  
 QY 1527 GVPDLPESPVPMRQ-----FVGCERNLOYDSR-----HI----- 1556  
 Db 3939 G-----HLRQGTGKRGAGVAGSFRGCMDSIYLANGELPLANKPRAVAHIEWV 3988  
 QY 1557 -----DMADFIANNGTV--PG-----CPAKNVCDGNTCHN 1585  
 Db 3989 DLSHGCLLTATBDCSSSPQNGGVNCPFTGYCYCKNALYVGTCEVSVNPPCSSNPCLS 4048  
 QY 1586 -----GGTCVNMQDAFSCPCPLGFGK 1607  
 Db 4049 GGTFTICDSGVCSGLVILCSDQLCSFCQDDPCQIGTGFPSDLDGAVCHDGSFRGE 4108  
 QY 1608 -----SCAQEN-----ANQHLFGSSLVAMHGLS 1631  
 Db 4109 RFQIDIDECAGNPCRNGLCEITLTSFYHCNCSQFGRKRCBDASPNHY----- 4156  
 QY 1632 LPISQFWYLSMFTROADGVLLQAITGRGRTITQLREGHVMSVEGTGLQASSLRLEP 1691  
 Db 4157 --VSTFPMNGL-----AEGIGIYFVINGIYLVNVVPLCRMIS-----RKKK 4197  
 QY 1692 GRANDQWHAQALASGGPCHALISPDYGOQRAEKNIG--PRLHGLHSNITVGIP 1748  
 Db 4198 ROAEPEB-----KRLGPTTALQRPYPFSKINKNIYFDIPAOVVRPISYTPSIP 4248  
 QY 1749 GPAGGVAAGFRGLQGVAVSDPREGVSLDP-----SHGESINVEGCSL-----PDPCS 1799  
 Db 4249 SDSRNNDL--RNSPFGSALPEHPE--FSTFNPSSMHGKRAVAV---CSAAPNLPPLPPPSN 4302  
 QY 1800 NPPCANYSYCSNDWDSYSCSDPGYGGDNCTVNDLNPCEHOSVCTRKSPAPH 1851  
 Db 4303 SPDSDSIQKPSWDP-----DYDAKVVDLDPCLSKKPLBEKPSQPY 4343  
  
 RESULT 8  
 Q8R508  
 ID Q8R508 PRELIMINARY; PRT; 4555 AA.  
 AC Q8R508;  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE FAT3.  
 GN FAT3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCB1\_TaxID=10116;  
 RN NCB1\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RT STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=21670969; PubMed=11811999;  
 RA Mitsui K., Nakajima D., Ohara O., Nakayama M.;  
 RT "Mammalian fat3: a large protein that contains multiple cadherin and  
 RT EGF-like motifs.";  
 RL Biochem. Biophys. Res. Commun. 290:1260-1266(2002).  
 DR EMBL; AB076401; BAB86869.1; -.  
 DR InterPro; IPR000152; Asx hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000742; EGF 2.  
 DR InterPro; IPR001881; EGF Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR001791; Lamthin G.  
 DR Pfam; PF00028; cadherin; 33.



DR Pfam: PF00008; EGF\_3.  
 DR Pfam: PF00054; laminin\_G\_1.  
 DR SMART: SM00112; CA; 32.  
 DR SMART: SM00179; EGF\_CA; 1.  
 DR SMART: SM00282; EGF\_CA; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL\_1.  
 DR PROSITE: PS00032; CADHERIN\_1\_19.  
 DR PROSITE: PS00268; CADHERIN\_2\_30.  
 DR PROSITE: PS00022; EGF\_1; 3.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR PROSITE: PS00025; LAM\_G\_DOMAIN; 1.  
 DR Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein.  
 SQ SEQUENCE 4555 AA; 502077 MW; B47CB010638BA73B CRC64;

Query Match 9.1%; Score 1413; DB 11; Length 4555;  
 Best local Similarity 26.1%; Pred. No. 2.7e-79;  
 Matches 520; Conservative 264; Mismatches 807; Indels 400; Gaps 72;

179 PQPQPSYQATYVENOPAGTPVASLRAPIDBGRAGRLBYMDALPDRSNOPPSLDPVT 238  
 2705 PPTQSQYSEFTLAEPTISIGSTIDTLRL--PQSVRFPTVNGERENKKNFTILEQST 2761

239 GAVTAEELDRKSTHVRVTRQDGMRRSLATLT---LVTDTDDHPVEEQEY 294  
 2762 GAIKDKRLDHEVSPAFHKVAT--IPLDKVDIVFTVDVVKVLDLNDKPEFTSSY 2818

295 KESLRLENLEVEYELVTRATDGDAPNPANILYRLLEGSG--GSPSEVFEIDPESGVRTKG 353  
 2819 ETLTMEGMVGTGLAQVRAIDTDGANGVYTSLSHSDHLKRYMEFNDISNTGMLSTLK 2878

354 PVDREEVESYQLTVEASDQGRDGPRTTAAPVLSVEDNDNAPQSEKRYVVOYREDVT 413  
 2879 DIDHETDPTFSEFVVASDLG-EAFSLSMALVSVKATDINDAPYFAHEVYNGVWESRP 2937

414 PGAPVLRVTRASDRDKS--NAVHYTSMGNARQGFILD--AQTGALDVSPLDYETTKEX 470  
 2938 PEEVAVASTLDKDTSNIRQVSYHITGNGPRGFALGMVQSEMKVYVRPLDREODIY 2997

471 TLRVRAQDGRPPLSNVGLVTVQVLDINDNAPIFSTPFOATVLESVPLGVLAVQAI 530  
 2998 FLNITRASDS---LFTQAMVEVTVSDVANDSPVCQVAYSLSLEDITSNKIILKVSXK 3053

531 DADAGDNARLEKRLAGVGHDPPTTNGTGMISVAELDREEDVYSPGEVARHDTPTAL 590  
 3054 DADIGSNGDIRYSLVSGNS--DFPLDPESEGLKTLALLDRERVPYNNIARATDGS--GR 3110

591 TIASASVSVTVLVNDNNPTFTQPEYVRLNEDAVGTSVTVYSAVDRDA--HSVITYQIT 648  
 3111 FCSSTVLLLEEDVNDNPVFSNNHYTACVYENTATKALLTRVOADVDPVGINRKVYSL- 3169

649 SGNTNRFSITQSOGGLVSLALPLDYKLERQVLAVTASD---GTRQDTAQIYVNVTD 704  
 3170 EBSAGSVBIDSS--GVIVLEQPLDREQSSYINISVRATDQSPGOSLSLTSVITVLD 3227

705 ANTHRPVQSSHYTVNVNEDRPAGTVVLISATDEDTGENARITFEMEDSIPO--FRIDA 762  
 3228 INDNPPVFERRDYLVTVPEDTSLGTQVLSVPATSDIGTNAITLILNSGNQGRINP 3287

763 DPGATVQAELEDEOVSTTALITARDNGIPQKSDTTTIELLVNDVNDNAPPLDSDYOG 822  
 3288 KTGGSIVLEALDYEMCKRYLLVEAAGGTPLSTAAVTSIDLTVNDNPPRFSQDVYA 3347

823 SYVEDVPPTVSLQISATDRDGLNGRVFTYFQGGDGDGDFIVESTGIVTLRLRRE 882  
 3348 VISEDALBEDSDYILLAEVDKSKNQIRFSIVG--DRNEPAPVDPIGLVAVKGLDRE 3406

883 NVAQVYLRAYAVDKMPAPRTMAEVTVTVLVNDNPPVFEODEFVVEENSPIGLAVAR 942  
 3407 RVSGVSLILQAVDSGIPAMSTTTVVIDISVDNDSPVFTPANTYAVIOENPVGSTSIQ 3466

943 VVATPDBEGTNA-QIMYQVBEGNIPVPLDIFSGELTALVDLDVEDREYVLYVQAIISA 1001

3467 LVVTRDRSEFHNPPPSFISLNGNEDEEFMLD--SHGILNSAVFRRHMESEYLLCIQAKDS 3525  
 1002 ---PLVSRATVAVRLDNDNDNPVLAGNEILLFNNTVTRSSFPFGAIGRVAPHDPDSD 1058  
 3526 GKPQGVSHTYIKRVRIESTHKTPTAIPLLEI---PIVWEDDFPGGVIGKHATQDDMD 3581

1059 SILVSPERGNELSLVTLNASTGELASRALDNNRPLEAMSVLVDGVSVTAQCALRYT 1118  
 3582 VLTFLAL-KSEQSLPLKVNASHGKIIALDGLDSGK---YLVANVSVDGRFQVIDVAVHE 3637

1119 IITDEMTHSITLRLIEDNSPERFISPLIGLPIQAVA---ATLATPDHVVENVQRT 1173  
 3638 QLVHEMLQNTVTVIRFENVSPEDF---VGLMHGFRRIILRNVALVLOKODSLRISIQ--- 3690

1174 DAPGHIINVLVSGQPPGPGGPPPLPSBDLQERLYNRL-LTALSAQVLPDPDNC 1232  
 3691 PVGVNQIDMLPAVEMHSSEYTKPAYTLQKLSNARHLENVHIALLEKNCGLD---C 3747

1233 LREPCENTMRCVSVLRFPSSAPFLASS--VLFRRPIHPVGLRCRCPPGFTDYCEYVDL 1291  
 3748 QEQHCE-----QGLSDSHALMTYSTARISFVCPRYRNRCTC---NGGVCEGSDP 3797

1292 CYSRPCGPHGCRSREGGYTCLCRDGYTGEHCEVSARSGRCTPGVCKNGGTCVNLVGF 1351  
 3798 CVKEKCPEDMOC-----VGY---EASRRP-----F 3819

1352 KDCDPSGDFEKPQCO-VTTRSPFASPTTR--GLQRHFTLTALSPATBEROGLALYN 1407  
 3820 LQCCPPGLGR--CSGHTSLSFAGNSYTKYLSSENSRE--DFKALRLRITQNSGIMY- 3875

1408 GFENEKDFVALLEVIOGV--QLTFSAGESITTVSPFVPGGVSDGOMHTVOLK----- 1458  
 3876 ---TRANCMILKIYEGGLMQLDQSGPGLIGIS---SRVNDGSMHSVLELNRTS 3929

1459 ---YNNK---PLIGCT----- 1468  
 3930 LSLDSYVERRRAPLYFQTLSTDSALFPGALVQADNIRSLTDRVTQVLGGFQGLDSV 3989

1469 ---GLP-----QGPSEOKV--AVVVPDGDGVALARFSV---LGANSQAA-- 1506  
 3990 LNNHLEPLQNRSSFAEVGTEKGLCVLPDQCSPLCHGSCSLPBGGYQSCLS 4049

1507 ---QGTGGSKSLDLTPGLLGGVPLDPESEPVEMRQFVGCERNIYDSSRHIDMADEFLAN 1564  
 4050 QFTGNCSEITACFPNCRNGSGCD---PIGNTFISCKAGL----- 4089

1565 NGTVPGCPAKKNVCDSNTCANGGTCVANOMDAFSCCEPLGFGKSCA--OEMANPOHFLGSS 1623  
 4090 --TGVTCEDVDDECEBCEBCEGSGSCVNLFGSPCNCTPGVYGCGLRPVVVPNIQAGHS 4147

1624 LVAMH---GLSLPISQPMYLSIMFRTQADGVLLQALTR---GRSTITLQAREHVMLSV 1677  
 4148 YVGBEELIGLAVLVFTLTLVLPF-----VPRKKYFRKNYSNNNTLLVODPATALLH 4201

1678 EGTGLQASSLRLBGRANDGDGMHQAOLAGSGGPGHAILSFYDGOQARABGNLGPRLHGL 1737  
 4202 KSNIGLPPRSLLA-----GGRNVYQEVGPQVPAVPMATPFCQDSDSRNLKGDAL 4254

1738 HLSNITVGGILGPRGGLARGRGLQGVRSVDPBEGVNSLDPSHGESINVEQG---CS-- 1792  
 4255 ---GG-----EPOELSTFEPESPRILITARGVVCSVA 4284

1793 ---LP--DPCDSNPPCANSYCSNDMSYS-----CSC----- 1819  
 4285 PNLRAVSPCRBD---CSIRKNGWDTGSEKKGADDTGEVTCFPAANSKNGSNEVSLANSFQ 4341

1820 ---DPGYGDNCTNVCNLPCEHSGVCTRKPS-----APH-GYTCBEPN-YLGPY 1865  
 4342 SDSGDNDAYHMD---TSDMMFGARLSDIBEMFYRESQDGAHVGQGSTRELSDYILGKY 4397

1866 ---CETRIDQPCRGWKGHPCTGPCNCVSKGFPDCKNTSGECCKE 1910

Db 4398 DISSETPPEHEEFLSDQQLPPPLP-----EDFPRQ 4428  
 Qy 1911 NMYRPSPTCLDCPTGSLSRVCDPBDGQCK-----FGV--ISGRQCDRC-DNP 1960  
 Db 4429 YEALPSPQSPTSL-----TGTMSPDCKRRPRFRHSQVLPPLPPLPGLTDLGGPSSCDPST 4482  
 Qy 1961 FAEVTNGCEV 1971  
 Db 4483 FAVSMNOGTEV 4493

RESULT 9  
 Q967F4 PRELIMINARY; PRT; 2920 AA.  
 ID Q967F4  
 AC Q967F4  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE HMR-1B protein precursor (W02B9.1b protein).  
 GN W02B9.1 OR HMR-1 OR W02B9.1B.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N2 Bristol;  
 RA Broadhead I.D., Pettitt J.;  
 RT "A Caenorhabditis elegans neuronal classic cadherin, HMR-1B, is  
 required for axon fasciculation."  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Lennard N.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY REPRESENTATIVELY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; AJ307058; CAC38842.1; -;  
 DR EMBL; Z82064; CAD27611.1; -;  
 DR EMBL; Z82093; CAD27611.1; JOINED.  
 DR EMBL; AL032638; CAD27611.1; JOINED.  
 DR EMBL; AL032638; CAD27619.1; -;  
 DR EMBL; Z82064; CAD27619.1; JOINED.  
 DR EMBL; Z82093; CAD27620.1; -;  
 DR EMBL; AL032638; CAD27620.1; JOINED.  
 DR EMBL; Z82064; CAD27620.1; JOINED.  
 DR Wormpep; W02B9.1b; CE30357.  
 DR InterPro; IPR000152; Aex\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR Pfam; PF00028; cadherin\_C\_term; 13.  
 DR Pfam; PF01049; cadherin\_C\_term; 1.  
 DR Pfam; PF00054; laminin\_G\_1.

DR PRINTS; PRO0205; CADHERIN.  
 DR SMART; SM00112; CA; 13.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00282; LamG; 1.  
 DR PROSITE; PS00010; ASK\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 8.  
 DR PROSITE; PS00268; CADHERIN\_2; 13.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00025; LamG\_DOMAIN; 1.  
 DR Calcium; Calcium-binding; Cell adhesion; EGF-like domain;  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 POTENTIAL.  
 SQ SEQUENCE 2920 AA; 323910 MW; 5B0D4C8BD24694FB CRC64;

Query Match 7.5%; Score 1164.5; DB 5; Length 2920;  
 Best Local Similarity 24.3%; Pred. No. 7.3e-64;  
 Matches 470; Conservative 279; Mismatches 745; Indels 437; Gaps 74;

Qy 176 NTAPOFOPSPYQATVPENQAGTPVASLRATIDPDGBAGRLLEYTMDALFDSRSG----- 230  
 Db 974 DNSPQERPSYETSVSFEANGTSVITVPAFNDAAHNAETYSLE--IDITAGEHQND 1031  
 Qy 231 --FFSL-DPVTAQVTAELDRETKSTHVFRTAODHGP--RRSALATLTIL----- 278  
 Db 1032 LDFEFLVNRSGEILILPIPKTKQ-PIFNVIADNDGIPALQSSAQVTLNVLDKQKA 1090  
 Qy 279 ----- 278  
 Db 1091 PKQTSPOCKPQITVDENVELNKILRCRAVSSGDSRNSDVIYKLTASGPGNKAKSKPR 1150  
 Qy 279 -----VTDTDHPV 288  
 Db 1151 QNKKEKNGEWEVYIMGLDYEQVNTTLITADMTSRVASTQTPVVEKVDVNDVVPQ 1210  
 Qy 289 FEQDEYKSL-----RENTLEV--GYEVLTVRA--TDGDAPNPANILYRLLEGSGSPSE 338  
 Db 1211 FTVDLFTGIDEMTPNEHLKTKNGKPIYTVVAIDTDSGPN--EWHRIYVEANGSEETK 1269  
 Qy 339 VEIDPRSGVIRTRPVRREAVESQVTVZASDQGRD-----GRSTTAANFLSVEEDN 393  
 Db 1270 HFRIDELTGEIIFPNKEFPREKIDMYILITVEASDRSVSLPGANGNKONKVKQIYINDN 1329  
 Qy 394 DNAPOSEKRYVQVREBDVTPGAPVLRVTAASDRDGSNAVVAYSIMSGNARQGYLDQT 453  
 Db 1330 DNAPSEKRYIGRYKESRGEHDVITIKAHLDHSLKRYTL-IGAGGRIPFGVRTDS 1388  
 Qy 454 GALDVSPLDVETTYEYTLRVARADQGRPPLSNVSGLVTVQVLDINDNAPFVSTPEQAT 513  
 Db 1389 GTIFVPEPLDFEASDQYHLVILASGRNATTN-----YIHIEDVNDNAPQEQOKYATT 1444  
 Qy 514 VL-ESVPLGLVLAHQALDADAGD-NARLEAYLAGVGHDPFTINNGCWISVALELRD 571  
 Db 1445 VIEEDVDLPKVLFNVAHTADQDERKSRIYVRLBQAGDEVRIKYSGLTELVRLALRD 1504  
 Qy 572 E--VDYFSFGVREADHGTALTAASVSVVTVLVNDNNPFTQ-----PEYTVRLNED 622  
 Db 1505 PRAGVPSNMFFVQALDDGNGLVGADVQVNVNRDINDNSPIPEHLFGVIEENRRPHSD 1564  
 Qy 623 AAVGTSVTVSAVDRAHSVITYQITSGTNR-----PSITSQGGGLVSLALPLDY 675  
 Db 1565 ---GVFMVDQARDDDPTTENANLEYGIVRKLTNGESVFRIDNTGKIFAMRSLDREI 1621  
 Qy 676 KLEROYVLAVTASD---GTRQDTAQIVNVNTDANTHRPVPQSHYTVANNEDRPAAGTVV 732  
 Db 1622 SSERETILVRANDGVPSSRSPFANVTIKVDMNNAPFEETIRYEGSVIEETAPLGAAMV 1681  
 Qy 733 LISATDEDTGENAR---ITYFMEDSIPOFRIDAD-----TGAVTQAELDYEDQVSYT 782  
 Db 1682 SFSAPFAD--EAKONVFYQLESBSDFYVYTTDQSKQSVGVLRVQKPLDYED----- 1734

```

QY 783 LATARDN---GIPKSDTTY-----LEILVNDVNDNAPQPLBDSYQSSVDEVPFTSV 834
DB 1735 --VTQBDGSHLAGI-RSDGRHDAEAVALVDRNHAPH-IHGATEHVRADVPFGTSTI 1790
QY 835 IQISATDRDSGLNGRVFTYFGDGDGDFIVESTSGIVRTLRDLRENAQVYLAAVAV 894
DB 1791 GRYTATDRDAGDTAR--FRINQSDPKRQFTIDQ-DGLRVAVHTDREDIAYNYLILEAY 1847
QY 895 DKGMPPARTPMEVTVLVLDVNDNPFVEDEDFDV---FVEENSPIG-LAAVAVTATDPD 949
DB 1848 DMSNNIGR--QWVAVVLQDVNDNGP---EPYVPPRCIFRRETPVNOQGTCEIRATDRD 1901
QY 950 EGTNAQIMVQIVEGNIPFVOLDIF-----SGELTALVDLDYE-DREYVLL--V 995
DB 1902 TAEFGPPFTMEVSPSKYSQYLAIVFNANGDGNKMTTLPQDFRPAVPGKILEPL 1961
QY 996 IQATSNAPLVSATVHVRLDRNDNPFVLGNFELFNNTYTNSSSPFGAIGRV--PAH 1052
DB 1962 ILADRAGRNRBASVHYIIGDLNDNTWHDGRMTIHNVSYGLRKET---VIGRVVVDAD 2017
QY 1053 DDDIDSLTYG-----FERGENELSVLL--NASTGELKLSPL--DNNRPLEAINSVLVS 1103
DB 2018 DWDLDGDK-TFSWKDSRPGFELSDKGSITWAGEMAAGTYMSANVHDNADDEDAVGVTVV- 2075
QY 1104 DGVHSVTAQACALRVITITDEMLTHSTITLEDMSPERFL-----SPILGLFIQVAAATL 1157
DB 2076 --IYVAVPOIAIDNGQSVOLLTAETPLQL-----PDDFIRADSNQSLMDTQKQEMTAYM 2129
QY 1158 ATPRHVVVFNVO-----RDTDAPGHIILANLSYGPGRPGGPPFLPSEDLQERLVL 1211
DB 2130 GGDVAV-VDFVSVOVGATLQTRDVP--VLAVRFN-----ARSTYRDTMQNLGLIA 2178
QY 1212 NRSLLTIAISQAVLPEDDNICLRPGCENTMRCVSLRPDSSAPFLASSVLFPRHPVG 1271
DB 2179 HRADIQKRLNVAIVGVGIDMCKFTQCD--AGQTLMSADYDGIIVVANSSTVIVGNATSR 2236
QY 1272 LRCRPPGFTGDCETEVDLCYSRPGPHGRKSRSEGGVTCGRDGYGHEHVSARSGR 1331
DB 2237 DDCTCP-----VBRAPPA 2249
QY 1332 CTPGVCKNGGTCVNLVVGFKCDSPSGDEKPYCQVTTSPAPHSPIFRGLRORFHTL 1391
DB 2250 CQHSICNDGVCHN-TNPGFPCRCNDGLKAGRCQOTTRSFGNGPAMWKMPACTSLINI 2308
QY 1392 ALSPATKRDGLLTVNGRF---NEKH---DPALEVYQEOVQULTFPAGESTTVSPPV 1443
DB 2309 SFSFMTTOSDALLFYNGPLETLRNTJHIEYSYIFIQLRGGRISLEVSNNQSGRSLEVA 2368
QY 1444 PGGVSDGQWHTVQLKXVNNKPLLGQGLPGSPSEOKVAVVTVYDGCCTGVALRP-GSVLGNY 1502
DB 2369 STALNDGTWHDISVN-----OEGKRVELVNDNC-----RFLGAGADDS 2406
QY 1503 SCMAOGTQSGSKSLDLTGPLLAGV---PDLSPESFVNRMRQFVGCNRLQVNSRHID 1557
DB 2407 SCRRLTYPPDDDERLNTVTPVQIGGLAPLSGGDYQTLF--RAGLNGCVCRNLNANGDYD 2464
QY 1558 MADFLANNGTVGCCTPAKKNVCSNCTCHANGCTCVNOMAFS-----CECPLEGCG 1606
DB 2465 LATPAFEONSEKGCYLMGATCDNSNSVDLSLNCIH-GDCEFADYQSGAMVAKVCPCPWGCG 2523
QY 1607 KSCAEMANPOHFLGSLVAMHGLSLPSQPYLSLMPFTRQADV---LQAIIRGRST 1663
DB 2524 ARCERRMEMIOPAQGAFLIYSRIAPF-EQVSDILFLPSGVNAPAPLBSGTDSSQSY 2582
QY 1664 ITLQIREGHVMSLV---GTG-----LQASSILRPERGRANDWMMHQAALGASGG 1711
DB 2583 VSTINLESQNGVTAAGKFDIGTGRRARQELRVSEVLL--KENASYM-----LQFTRN 2633
QY 1712 BEHALTSPD---YQOQABEGMLGPRHGLHSNTTVGGIIPBPAGVARGPGCL----- 1762
DB 2634 PFRASLSDINAYTVSTQIDRG---EPFELQVNOQITLG-----TOGQNKFGQCGCTGYRW 2684
QY 1763 -----QGVAVSDTPEGVNLSIDPSHSGSINVEGCSLPDPCBNP---CRANSYCSND 1811

```

```

DB 2685 SKONIPLRKGAMDNESIVSISNMAG-----VQDGDRLITTCADLPAGYCGGSPVQVDF 2740
QY 1812 WDSYSCSCDPG 1822
DB 2741 WKGPCTCNDG 2751

RESULT 10
Q8MY75
ID Q8MY75 PRELIMINARY, PRT, 1959 AA.
AC Q8MY75;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Pfl-cadherin.
GN PFLCAD.
OS Ptychodera flava.
OC Eukaryota; Metazoa; Hemichordata; Enteropneusta; Ptychodoridae;
OC Ptychodera.
OX NCBI_TaxID=63121;
RN (1)
RP SEQUENCE FROM N.A.
RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Sato N., Humphreys T.,
RT Zhang S., Tsukita S.;
RT "A novel amphioxus cadherin that localizes to epithelial adherens
RT junctions has an unusual domain organization with implications for
RT chordate phylogeny."
RT Evol. Dev. 0:0-0(2002).
CC - FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY REPRESENTATIVELY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AB075368; BAC06837.1; -.
DR InterPro: IPR000152; Asx hydroxyl.
DR InterPro: IPR002126; Cadherin.
DR InterPro: IPR000233; Cadherin_C-term.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR006209; EGF_1like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00028; cadherin; 5.
DR Pfam: PF01043; cadherin_C-term; 1.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00054; laminin_G; 1.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 8.
DR SMART: SM00181; EGF; 3.
DR SMART: SM00282; Lams; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00232; CADHERIN_1; 5.
DR PROSITE: PS50268; CADHERIN_2; 8.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS50025; LAM_G_DOMAIN; 2.
KW Calcium; Calcium-binding; Cell adhesion; EGF-like domain;
KW Glycoprotein.
SQ SEQUENCE 1959 AA; 214791 MW; F778D75002F084CF CRC64;

Query Match 7.1%; Score 1108.5; DB 5; Length 1959;
Best local similarity 24.0%; Pred. No. 1.3e-60;
Matches 445; Conservative 252; Mismatches 703; Indels 455; Gaps 72;

QY 176 NTAPQFPQPSYQATVPENOPAGTFPVASLRAIDPDEGAGRLGYTMALFDSRSNQPSLD 235
DB 216 DNABEPPEPTECEVTVPEISPDALVAIVHAVDPDAGKNGKMYVYTD-----NDMTSVH 270
QY 236 PVTGAVTAAEHLDEBETKSTHVFVTAQDHGMPRSSALATLTITVDT----- 282
DB 271 PDTGVIIGGFLQAKREYLVF---ARDRGKPSLEASTPGVITVTPVDSILKSSILG 327
QY 283 -----NDHDPVPEQGGYKSLRENLEVGVEVLTVRAITDGDAPANNAILVRLLEGSGGSP 336

```

```

Db      328 PGVILLDDDE--ELPEIKVSLMEDAPKGSVTTITKPSNDKSSQ-----RFLMNNMNP 381
Qy      337 SEVEIDPRSGVIRTRGPVDRREEVSYQLTYEASDQGRD---GPRSTAAVFLSYEDDN 393
Db      382 R--FHVQYVGDVILTRLDIDAHPTIEINVGITKDNDRPLRBPQKQKIIT 1PSET 438
Qy      394 D-----NAPOFSEKRYVVOREDVTPGAP---VLRTASDDKGSNAV 433
Db      439 DKPAPNQVVKPRIRRAKSPSFDQPSYEGTVEVDGLASPVPAVPRPILITVANGPSPN 488
Qy      434 VHSVIMSGNARGQFYLLAAGALDVASPLDETTK-----EYTLVRNRQDGRPLASVSG 489
Db      499 IORHIOGAERREDPAIGTSGIISRKHLREDNKGQVREFTVA-----SSAG 548
Qy      490 LVTVOVL-----DINDNAPIFVSPFOATVLESVPLGLVLAHQALDAD---AGDNALLEY 542
Db      549 TTSVEVLKIKDVNDNAPQPPYPYISPPENGAPQVTVMTIQEDLDDBNEGNNLVRY 608
Qy      543 RLA-----GVGHDPFTINNGTMTISV-AAELDEEVDFTSPGYEARDHGTPLTASAV 596
Db      609 SIIRMAONSVOAL-PAIDSESAVITTOVGNDREELPSYDIYKAVDGG--AHTGVTA 665
Qy      597 SVTVLDVNDNPPFTQEPYTRLMEDAAVGSVTVASVDRDANSVITTOITSGENTRF 656
Db      666 TIYVSDNDNPPPEPOAKSFVEISENPVDSVFKVLAIDRV-GINSYISISGNVEKF 724
Qy      657 SITS--OSGGGLVSLALPLDYKL-EROYVLAVTSADGTRODQAIVNVNTDANTHRPVO 713
Db      725 RIETDPDSGALVAVAKKIDPELINESKVLITIBVDSKNTATTELDIRVANDANEVSPS 784
Qy      714 SSHYTVANVEDRPGCTTVLISATDEPTGEN-----ARITYMEDSIPQFRIADTG 765
Db      785 QPRVLAHPEDPAKMTTIGVAAATDRKSEKSEIEBEVAVRPFPPSSR--FVVGANG 842
Qy      766 AVTQAEID---YED---QVSYTLATARDN-----GIPOKSTTYLILVNDVNNAP 813
Db      843 SWAIKULDERPYEDDPOIRENYITVYATIDNVHGGGERANTVTTFTVISITDNNAP 902
Qy      814 QPLADSQGSVYEDVPPFTSVLQISATDRSGLNGRVYTFQGGDDGDFIVESTSGIV 873
Db      903 EFL-PGYTLVPEMTPGAYEAGVVEIIRDDDSNGPP----- 939
Qy      874 RILRLRENVAAQVYLAIVAVDKGMPAKTMEYTVTVLVNDNPPVPEDEPVPVEEN 933
Db      940 -THRVIDDGILSR-----PDV---S 956
Qy      934 SPGLAVARVATADPDREGTNAQIMYQIYEGNIPEVFOLDIPSGELTALVDLYEDRDEYV 993
Db      957 SPIC-----VEDQI-RVSLLEEDREAITPPY-----YI 983
Qy      994 LVIGATSAPLVSRATVHVRLLDRNDNPVLGNFELFNNTVYNNSSSPGGAIGRVPAHD 1053
Db      984 PVLVLSDKHGVSATQTLVTIGDVNDNPPSPATKITLVYSY---KGGIPRAIGEVYAAD 1039
Qy      1054 PDI--SPSLTY-----SPEKGNELSLVLAASSTGELKLSALDNNRLEALIMVLYV 1102
Db      1040 EDIVKGDISTYVTPPIPOFVSVDKDTQGIIEBTKAGTYDKVKEVDEKPPVICVYIV 1099
Qy      1103 SDGVHSTVAQCALRTVITTDMLTHSTLRLEDSPERFL---SPLLGLFIOAVAAATLAT 1159
Db      1100 -----EYKDIPEBAVPSGSSVRL-LSAEKFTINGTSMKDRFDILVDVGA 1145
Qy      1160 PRDHVVENVORDTADPGHILINVSLSVGQPPG-----PGGPPFLSEDLQERLY 1210
Db      1146 KKENIDIPSY-----INV-----PGEENQDVMSAHGSPYVPEKMTQTL 1187
Qy      1211 LNRSLTLTISAQVLPF-----DNICLARECEMYMACVSLARDSAPRTASSSV 1261
Db      1188 LAKDV---IESQLGVSFGKVPIDMCLDRGIC-ESSCTNYFNAI-----TDPITIDGAD 1237
Qy      1262 LFRPIHVGGLRCRC-----PGFPGDYCETEVLDLYSRPGCPHRCRSGGYTCLCRD 1316

```

```

Db      1238 TVSVTTIORACVCGAKIAPRG-----PCASIPCLANGGOCIDTBSGYCCQDE 1286
Qy      1317 GYTGHECEVSARSGRCTPGVCNNGTCVNLVGVFKDCPSGDPEKPYCQVTRSPAS 1376
Db      1287 KCGG-----PNCEDISRGQNS 1304
Qy      1377 FTTFRGLRQRFHTLALSPATKERDGLLYNG-----RNEKHDVPALEVIOEQVLTSS 1431
Db      1305 FAWYPLSQCEBTKTISIEFLTSSRDGIILYNPIVPIIDLGQPDFWATELVGKPELXMD 1364
Qy      1432 AGESTTVPFPGVSGSDGQMTVOLKYNRKFLGQTLPGQPSRQKQVAVTVVDCDTGV 1491
Db      1365 LSGMLPLTIDSP-PLNDKMTLTI-----IRSKHVEPLDHCENAV 1407
Qy      1492 ALRF--GSVLGNVSCAAQGTQGSKKSLDTLTPLLLGVP-DLPSFPVRM-----QPV 1543
Db      1408 IYETERSTVSDSNCAATDPTTGTRFLVNTPLQGGIDMDSTYTPYTFNPAFGDVF 1467
Qy      1544 GCMRNLOVDSKRIIDMADFLANNQVPGCAKQVCDN---TCHNGCTVNMOWAFSCPC 1600
Db      1468 GCIRVYIDGKRYDLDGTPGREKNSQGCPTDLCTVDGVYCYKGLCVTESASVYTCV 1527
Qy      1601 PLGFGKSCAOMANPOHFLGSSLVAMHGLSLPISQP---YLSLMTFRQADGVV--LQ 1655
Db      1528 DEGWGEACDEBLPS-YDPLSDSYVYIDLAASLVNDKMTSDQIMVGNTOBEDALLMHQ 1586
Qy      1656 AITRGSTITTLQLRBGHVLAYE-GTGLQASSLRLEPGANDGWHHAQALGASGPGH 1714
Db      1587 NVD-GEYITMKGNAFLIENYLDGPMKYSLIRVANYTINDGNVTVL----- 1635
Qy      1715 ALLSPYQQRABGNLPGRLHGLHSNTVGGIIPGAGVAVAGFRGCLQGVSDPREGV 1774
Db      1636 -----LYGNIFRFEDG--GG--GVTRVKAQGI 1661
Qy      1775 NS---LDPFH---GESINVEQCSLPDPCSNPCPANSYCSNDMDYSQSCD-----PGY 1823
Db      1662 FKRPIDPSVYGVANVSLSERVLGD--DYGCACADAISNSYISFDEIDISTIAVPR 1719
Qy      1824 YGDNCTNVCDLNPCEHOSVCTRKPSAPHGYTCBPEN-YLGPYCETRIDQPCR 1876
Db      1720 VERDC-NICLSFCLAPLVCPPTPSF---YCSCPEDLVFRGAC-VKGGPAPQ 1769

RESULT 11
ID Q19319 PRELIMINARY; PRT; 4307 AA.
AC Q19319; Q19785; Q21606;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical 479.0 kDa protein F25F2.2 in chromosome III.
GN F25F2.2.
OS Caenorhabditis elegans.
OC Bkaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Ainscough R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Z35662; CAAB4721.1; -.
DR EMBL; Z35599; CAAB4721.1; JOINED.
DR EMBL; Z34802; CAAB4721.1; -.
DR EMBL; Z35599; CAAB4661.1; -.
DR EMBL; Z35662; CAAB4661.1; JOINED.
DR EMBL; Z34802; CAAB4661.1; JOINED.
DR EMBL; Z34802; CAAB4339.1; -.
DR EMBL; Z35599; CAAB4339.1; JOINED.
DR EMBL; Z35662; CAAB4339.1; JOINED.
DR HSP; P00740; IEDM.
DR WormPep; F25F2.2; CE00961.

```



Db	3949	-----E 3949
Qy	1796	PCDSNPGCANVCSNDMDNSYSCSDPGYYGNC-----TNYCDNPGCHSVCTRK 1846
Db	3950	PCASSPCTGTGTCQCPFFINDYLCKBNGFTGKHCARGFEDHETSSCSKNVCGTSGCQICSI 4009
Qy	1847	PS---APHGYTCEC 1857
Db	4010	FRHSLESDFCINC 4023
RESULT 12		
ID	Q8N5B3	PRELIMINARY; PRT; 1061 AA.
AC	Q8N5B3;	
DT	01-OCT-2002	(TREMBLrel. 22, Created)
DT	01-OCT-2002	(TREMBLrel. 22, Last sequence update)
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)
DE	Similar to cadherin related 23.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Eye;	
RL	Strasbourg R.;	
RU	Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.	
DR	EMBL; BC032581; AAB32581.1; -	
DR	InterPro; IPR002126; Cadherin.	
DR	InterPro; IPR002114; HP_Serp_site.	
DR	Pfam; PF00028; cadherin; 9.	
DR	PRINTS; PR00205; CADHERIN.	
DR	SMART; SM00112; CA; 9.	
DR	PROSITE; PS00232; CADHERIN_1; 6.	
DR	PROSITE; PS00268; CADHERIN_2; 7.	
DR	PROSITE; PS00589; PTS_HPR_SER; 1.	
KM	Calcium; Calcium-binding; Cell adhesion; Glycoprotein.	
KQ	SEQUENCE 1061 AA; 11611 MW; A97PA5F07EFCB935 CRC64;	
Query Match 7.0%; Score 1081.5; DB 4; Length 1061;		
Best Local Similarity 30.8%; Pred. No. 2,3e-59;		
Matches 340; Conservative 159; Mismatches 457; Indels 147; Gaps 36;		
Qy	175	VNTAPQPPSPYQA--TPENQAPCTPVASIRALDPDE----GAGRLRYTMALPFSR 227
Db	25	VNRLPEFFNMFDDYLLISBDTPVGSSTYOLLAQMDMDPLVFGVS-----EE 73
Qy	228	SNQFSLDPVGVATVTAELRLRETSTVFRTYADH-GMPRSALATLTLLTVDNDHD 286
Db	74	ASRFFAVSPDGVVWLRQPLDRBKTSPTVSPSVSDHGVTRK---VNIQGVANDNA 129
Qy	287	PVPEQEEKESLRNLRVGYEVLTVRATDGDAPPNANILLYRLBGSGSPSEVEIDRS 346
Db	130	PTFNHQPSVRIPEPTPVGTPIFIYNAADDPDLAGGSVLY-----SFQPSQFALDSAR 184
Qy	347	GVIRTRGVDEEYVESYQLTVEASQGRDPPGRSTTAIVPLSVEDNDNAQPSSEKYV 406
Db	185	GIIVTVIRELDVETQAYQLTVNATDQK-TRPLSTLANLAILTDVQMDPIFNLPEST 243
Qy	407	QVRBDVTGAVLRLTASDRDKGSNAVVHYSIMSGNARQGYLDAQTGALVNSPLDET 466
Db	244	NIYHSPPTGTVIRITTAIDQKGRRGIGYTLVSGNTSIFALDVISGLVTLGLNDREN 303
Qy	467	---TKSYTLRPA---QDGRPLSNVSGLYTVQVLDINDNAPFVSTTPQATVLESVPL 520
Db	304	PLYSHGFLTVKYGELINDRTPSDATVTTTFNIIVIDINDNAPENSSEYVALTELAQV 363
Qy	521	GY-LVLAHQALDAD--AGDNARLEKRLAG-VGHDPFTINNGT---WISVAELDRE 572
Db	364	GFALPLPIQVWDKDKENLGNSFEYVLYVGNNSHHIIFSPVQGGADIRIRVAPLDVET 423
Qy	573	VDYFSFGVABDHGTPLTASASVSVTLVDVNNNPTFTQPEYTVRLMEDAVGTSVTV 632

Db	424	VDRDPLFLPANE-SVPHGVGA	AKITLLINENDNRPIFSQGLYNI	SLYENTVGGISVLTV	482		
Qy	633	SAVDRDAHSVTTV	ITSGNTRNRSSITSGSGGLVSL	ALPLDYGLEROYLA	VTASDQTR 692		
Db	483	LATNDGATGCEVSV	PFSSDDPDRSLKDT--GLIM	IARLDYELIQRFLLTII	ARDGG 540		
Qy	693	QD-TAQIVVNTD	ANTRPVQSSHXYTVNV	EDRPAGTIVVLISAT	DEDTGENARITYFM 751		
Db	541	EETGRIRINVL	DVNDNVPTFGQADYV	GALEKRENPSTQVLR	LATDEDSPPNNQITYSI 600		
Qy	752	ED----	SIPFRIDADTGA	VTQAELDYEDQVS--YT	LATKNGCIPOKSDTTYLEI 803		
Db	601	VSASAFSGSYED	ISLYEGEYVSVSRPLDYE-Q	LSNGLIYLVVMMADGN	PLNSTVPTI 659		
Qy	804	LVDNVNNAPO	FLDSYSGSVYEDVPPFTSV	LOISATDRDGL--NGR	VYTFQGGDDG 860		
Db	660	EVPENDNPP	FFSKAPAFVSVEEM	AGATVLFINATDLD	RSREYQGESIYSLEGST-- 717		
Qy	861	DGDFIVESTG	IVTTLRLDRE	NAQVYLRAYADK	MPARP--MEVTVIVLVNDP 918		
Db	718	--QRI	NARSGBITTTSLD	RETSEYILLVRAVDG	VGHQKGTGIATVINITL	DINDNH 775	
Qy	919	PVEQDEPDPV	VEBNSPGLAVARV	ATDTPDEGNAQIMQI	YVGNIPVYQDLIFGSL 978		
Db	776	PTWKADYV	INLVMTPPSDV	TVTVAVDPLG	ENGTVLSIOPPN--KFS	LSNSTYTKI 833	
Qy	979	T--ALVD	LYEDPEPEYL-----	VIQATSA--PL--	VSRATVVRLLDR	NDNPVLGNF 1026E	
Db	834	RTTHAMLD	RENPDDEHLE	MRKIVSVSTDCGR	PLKATTSATVYNLLD	INDNPPTQNL 893	
Qy	1027	EILENNYV	NRSSSPGCG-AIGR	VPADPD--ISDS	LTFSERGENEISLV	LNASTGELK 10833	
Db	894	P----	FVAELYEGIP	AGVSIYQVADL	DEGLNGVLSYRPM	GVMPMDPLINSSGVVV 948	
Qy	1084	LSRALD	NNRPLEALIM	SVLSDGVHVSVA	QCALRTIITD	EMLTGSIITLR	EDMSPERPLS 11434
Db	949	TTTELDER	RLAEYQIR	VASDAGCTPTSS-----	TTTLTHIVLDVND	E----- 991	
Qy	1144	PLIGLFI	QAVAAATL	PPDPDHVVENV	VQDTPAD-----	GGHILNVLSI-- 1186	
Db	992	-----	TTTPP	AVYVNVSESDP	REFRVWMLCTD	NDVG--LNAELSYF 10333	
Qy	1187	-VQ	PPG-----	PGCGPPPLP	1201		
Db	1034	ITGA	PASAHLCRPP	GALPPPLP	1056		
RESULT 13							
Q8T5J8 PRELIMINARY; PRT; 1881 AA.							
AC	Q8T5J8						
DT	01-JUN-2002	(T	REMBLrel. 21, Created)				
DT	01-JUN-2002	(T	REMBLrel. 21, Last sequence update)				
DT	01-MAR-2003	(T	REMBLrel. 23, Last annotation update)				
GN	11N17.9.						
OS	Anopheles gambiae	(African malaria mosquito).					
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;						
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicidae; Anopheles.						
OX	NCBI_TaxId:1165;						
RM	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=PEST;						
RA	Thomasova D., Ton I.Q., Collins F.H., Kafatos F.C.;						
RL	"Sequencing and analysis of Feni region from Anopheles gambiae						
RL	chromosome 2R.";						
RT	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AJ439060; CAD2762.1; -						
DR	InterPro; IPR002126; Cadherin.						
DR	Pfam; PF000028; cadherin; 13.						
DR	SMART; SM00112; CA; 13						
DR	PROSITE; PS00232; CADHERIN_1; 9.						

DR PROSITE; PS50268; CADHERIN 2; 13.  
KW Calcium-binding; Cell adhesion; Glycoprotein.  
SQ SEQUENCE 1881 AA; 209226 MW; 109F3F517BAF414 CRC64;

Query Match 5.8%; Score 906; DB 5; Length 1881;

Best Local Similarity 26.0%; Pred. No. 8.5e-48;

Matches 315; Conservative 176; Mismatches 413; Indels 308; Gaps 42;

```

QY 182 QPESYQATVP-----ENQACTPV-----ASLRALDPDEGEAGRLTYTMDALFDSRS 228
DB 275 QPPEFLYQPVVRISEDAPICTEVLARMIVSLSTKAVDGRGINNRIYGI-----SNGG 330
QY 229 NQPSLDPVGTAVTAELDRETKS-----THVRVTAQDHG--MPRSALATITLIVTD 281
DB 331 SELEIDRLKGLSTKQKLDREDSNTPNGAFILAEVIESKIQAPASSTMEIYVTD 390
QY 282 TNDHDVPEEQEYKESLENTLEVGVLTVRATDGDAPPNAILYRLLEGSGS----- 335
DB 391 VNDSEIPIPRSDGVCGEIENAG-----ENTLAFIDG-----SINEVPDIDQKNGTFRSLH 443
QY 336 -PSEVFEIDPRSGV-----IRTRGP--VDREEVSYQLTVEASDQGRDPGRSTTAVF 386
DB 444 PPSDIFFEVIFPKALINEATFGLRVKDPMDYERVELSTVVASGV--ESAGRTSTQIR 501
QY 387 LSVEDDNDNAPQSEKRYVQVREDYTPGAVLAVTASDRKQ--SNAVHTYSIMSGAR 444
DB 502 VVAVLDQNDNPFEPQPYVDIDVPRNVAIGVLLQIQATDSGSLFTEGVRYANILTGIS 561
QY 445 GQPYLDAQTGALDVV---SPL-DYETTKETYLRAQDQGRPLSNVSGLVTVQVLDND 500
DB 562 SFHLHDHAGVITLMASBSVYFDEIITQKHLSTSEADNGRGRNTVPLI-LAVLDVND 620
QY 501 NAPIFVSTPPQATVLESV---PLGYLVTHVQALDADGNALRYRLAGVGDHPPTI 555
DB 621 NAPIFVDRKREYRLEKNAFEFESP---IVPEARSDLEGSPNSAVERRLGASHSDYFH 677
QY 556 NNGGWTISVAALDREEDF-----YSGVETARQDGTALTRASVSTVL 601
DB 678 DRRGTGRISV---REANDFERLSSGSGSDRTIALITTEADGSEPLTAQVEVTVYQ 732
QY 602 DVANDNPFQTPQETVRLNEDDAVGTSVTVSAVDRDA---HSVITYQITSGNTRNFSI 658
DB 733 DVANDYAVFLSQAIVIPEDTPSGLVLRVLTAMDGSGSPNNHVTTRIQGGD-GAFVI 791
QY 659 TSGSGGLVSLALPLDYKL-----EROYVLAVTASDGTROD---TAQIVNVTDANT 707
DB 792 GASTGEISITHGASLIDPNLLAPDALGSGSFYLEVPAIDGNGDQQLQSGCLVNT- 847
QY 708 HRPVFGSHYTVVNNBRPAGTIVVLSATDEDEGENARITTFMEDSIPQ----- 757
DB 848 -----PVGETEYRLMATDPD--EGAMLRYYIDRSLSGKTEBEGALVK 887
QY 758 -----FRIDADGAVTTQAELEDYEDQVSYTLATARDNGIPOSDD---TLYLEILV 805
DB 888 LDDYDFAAALINBTNLKIAKLDRKIAETILACVDEVA-AEKGDDQMANITFLKIV 946
QY 806 NDVNDNAPQPLRDSYQSGSYVEDVPPFTSVLQISATDSDGLNGRVFTFOGQ----- 857
DB 947 LDEDDNPKFRKPKPYKHSIAENSQYGAVCTVVAEDADQ--NKTVKXSLBEGKGVLELH 1004
QY 858 -DDGDDGFIYESISGIVRTLRILDRBNVAQVLRAYAVDKGMPARIPMEVTVTVLDND 916
DB 1005 VDDGTGIVVRN-----RIDHEYSMLNPSVRADTGTPEPRASFVEVFQVQDEND 1055
QY 917 NPPEFQDEDFDVEENSPIGLAVARTATDPDEGTAQIMYQIVGENIEVEFQOLDIFSG 976
DB 1056 NNPFYVDSVNDYVYSENASGABAIILADLDOSGDGRITTYILDRSSSEKXSIDBERG 1115
QY 977 ELRALVDLYEDREPVYLVIOA-----TSAPLVSRAVTHRLDRNDNPPVL- 1023
DB 1116 ILKRVAGALDREETAELVMAVEAMDNYPGYLNGESRRAAFKHLIIV--LDDNDNVPIQK 1173
QY 1024 -----GNFEILFN----- 1031

```

```

DB 1174 PSCGSMTEBYNNINDPIVKLATDADPTNGNQLSFDIYDPGFIYIQVSAQYAEIYS 1233
QY 1032 -----NY-----VNNRSSSP----- 1042
DB 1234 RGPLKHLNGVITYLELVSDIGVPTNARBSVDICVDENHAFVFPVPSGNTVYKVENT 1293
QY 1043 --GGAIGRVPAHDPDISDLSYSE-----RGNELSLVILNASTGBIKLSRALDNR- 1092
DB 1294 TLGKPFQVAVYDEVDGENNAIVRYRLKMDTMGN-PRFSLDKETGELSLAFLDREQMM 1352
QY 1093 -----PLEALMSVLY-SDGHSVTAQCALRYTII--TDEMLTHSITLLED 1135
DB 1353 YDLRIEAYDQGIPTPLSTVDLIYVADVVDNLPQFLIKEISINTEHMTPTGRIERLPD 1412
QY 1136 MSPERFLSPLLG 1147
DB 1413 TVDDQYLDPLDG 1424

```

## RESULT 14

```

096JL3
ID 096JL3 PRELIMINARY; PRT; 803 AA.
AC 096JL3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA1812 (Fragment).
GN KIAA1812
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 1 [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Ref. 8:85-95(2001).
DR EMBL; AB058715; BAB47441.1; -.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 6.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS50268; CADHERIN_2; 5.
KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;
KW Glycoprotein.
FT NON TER
SQ SEQUENCE 803 AA; 88205 MW; 15F2EDF008FFAB9 CRC64;

Query Match 5.8%; Score 905.5; DB 4; Length 803;
Best Local Similarity 35.0%; Pred. No. 2e-48;
Matches 254; Conservative 104; Mismatches 326; Indels 41; Gaps 15;

```



QY 476 AODGRRPLSNVGLVTVQVLIDINDNAPFVSTFQATVLESYPLGLVLAHQADADAG 535  
 234 AMAGNPNPL-NSITVPVTEVFENDENDNPTFSKAPFVSVENIMAGTVFLNAITDLS 292  
 QY 536 DNARLERLAGVGHDPFTTNGTGWISVAELDREVDVPSGVEARDHGT--PALTAS 593  
 293 REYGOESIIYLSGASTQFRINASGEITTTSLIDREKSEYIILVRAVDGVEGHNOKTGI 352  
 QY 594 ASVSVTVLVNDNNPFTQPEYTRNLNEDNAGSVTVASVARD--ASHVITTYQITSGN 651  
 353 ATNITITLIDNDNHPFKADAPYITNINLVEMTPPSDVTTVVAAPDLAGENGLTVSIOPPN 412  
 QY 652 TRNRFSITSGSGGLVSLAL-----PLDYKLERQVYLAVTASDGT-----QDTAQIV 699  
 413 --KFYSINSTTKRIKRTTHAMLDRENPDPEHAEIMRKILVSVT--DCGRPLKATTSATVP 468  
 QY 700 VNTDANTHPVQSSHYTVANVEDRAGTIVVLISATBEDTGNNARITTFMEDSLPQ-- 757  
 469 VNLIDLDNDPFTQNLPEFAVEVEGLEIPAGVSIQVVAIDDBEGINGLVSRYMPGVGRMD 528  
 QY 758 FRIDADTAVTQALDYEDDOVSYTALATARDNGIPKSDPTYLEILVNDVNDNAPQFLR 817  
 529 FLINSSGGVVTTEHLDREKIAEYQAVASDAGTPTKSTITLITVLDVNDDETPTFP 588  
 QY 818 DSYQSVYEDVPPTSVLQISATDRDGLNGRVFYTFQGGDDGDPVSTSGIVRTLR 877  
 589 AVTVSVSEVDVPEFRFVVMINCNDVGNALSLFYITGG--NVDGKFSVGRDAVVRVTV 647  
 QY 878 RLRENNAOVTLAAVANDKMPARTMET--VTUVDVNDNPPVQEDFVQVFEENSP 935  
 648 GLRETTAAVMLILEALDNGPVGRKRTGTATVPTVLDVNDNRPITFQSSYEASVPEIDP 707  
 QY 936 IGLAV 940  
 Db 708 EGHSI 712  
 RESULT 15  
 Q9VVG0 PRELIMINARY; PRT; 1820 AA.  
 AC Q9VVG0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Putative cadherin precursor (CG6445 protein).  
 GN CG6445.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=2019606; Pubmed=10731132;  
 RA Adams M.D., Ceiliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Aoril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkov A.D., Bortolin M.R., Bouck J., Brooksstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Paolo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Garb M.W., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li U., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Maesarm D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL ADHESION.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 DR EMBL: AE003524; AAF9351.1; -.  
 DR HSSP: P15116; INCU.  
 DR FlyBase: FBgn0036715; Cad74A.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00208; cadherin. 12.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA, 13.  
 DR PROSITE: PS00232; CADHERIN 1; 4.  
 DR PROSITE: PS0268; CADHERIN 2; 13.  
 DR KMW Hypothetical protein; Cell adhesion; Glycoprotein; Repeat; Signal;  
 KM Calcium-binding.  
 FT SIGNAL 1 14  
 FT CHAIN 15 1820 POTENTIAL.  
 FT DOMAIN 36 128 CADHERIN 1.  
 FT DOMAIN 137 245 CADHERIN 2.  
 FT DOMAIN 254 365 CADHERIN 3.  
 FT DOMAIN 374 483 CADHERIN 4.  
 FT DOMAIN 492 585 CADHERIN 5.  
 FT DOMAIN 590 696 CADHERIN 6.  
 FT DOMAIN 718 815 CADHERIN 7.  
 FT DOMAIN 824 939 CADHERIN 8.  
 FT DOMAIN 955 1053 CADHERIN 9.  
 FT DOMAIN 1062 1170 CADHERIN 10.  
 FT DOMAIN 1170 1284 CADHERIN 11.  
 FT DOMAIN 1295 1394 CADHERIN 12.  
 FT DOMAIN 1407 1517 CADHERIN 13.  
 FT CARBOHYD 191 191 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 371 371 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 678 678 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 727 727 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 868 868 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 890 890 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1111 1111 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1112 1112 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1233 1233 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1294 1294 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1419 1419 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1452 1452 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1557 1557 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1584 1584 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 FT CARBOHYD 1610 1610 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SQ SEQUENCE 1820 AA; 200096 MW; 8D61B907FA15C155 CRC64;  
 Query Match 5.8%; Score 905; DB 5; Length 1820;  
 Best Local Similarity 25.4%; Pred. No. 9,3e-48;  
 Matches 350; Conservative 214; Mismatches 468; Indels 346; Gaps 55;  
 QY 180 GQGP-----SVQATVPENQPGTGVASRALDPDGEGRLELYTMDLFDLSRNSQFSL 234

```

Db      | 239 QDDPRTTNAFYSKIVENTPAGSIIITVAVNDGVDIIPRIRIFLSLE---DEPPGHELV 235
Qy      | 235 ---DEVTGAV---TTAEILDRETK-----STHVRVTAQ---DHGMPRSALATITLV 279
Db      | 296 PFGPGRGTAVLQTTSEPLDRENAEILONGVYFYSIRATELIDGALPAHSLTRVITV 355
Qy      | 280 TDNDNDHPEVQEQEKESLRENLVQ-----305
Db      | 356 TDVDHQPFTSGPHFNVSTIENLANGMPPLGLSI FVDDRDGMGNSRYELSLRDVFNARV 415
Qy      | 306 -----YELT-----310
Db      | 416 FEVSPTESQGRTPVVVVLNASRLDYVDPDLKPEFDLVASVKVAKTVEIHLTD 475
Qy      | 311 -----VRATDGAAPPNANILYRLLEGSGSPSEVFE 341
Db      | 476 ANDNAPVDOGTYRFTAEMLPVDAIIGHYKATDLSGEPGHVRY-VLKQFG---ADNFY 531
Qy      | 342 IDPRSGVIRTRGVDEBEVESYQLTVEASDQGRDPGRSTTAAVFLSVEDNDNAPQFSE 401
Db      | 532 VNPETGCVIILKPLDYEKSSYSTVVALDQO---REANANILFVGVTVDVNDHNPFS 587
Qy      | 402 KRYVQVRBEDVTGAPVLATVTAASDRDKS---NAVHYSIMSGN---ARGPYLDAQGALD 457
Db      | 588 KEYSRTIREGALPEPOFVRADADQPSQNGRVKKSIVSENSIAGNVFRIEPTGEIV 647
Qy      | 458 V---VSPLDYETTKETTLRAQDGGRPPLSNVSGLVTVQVLDIN-----DNA 502
Db      | 648 IQAKASMDTE-RGEYELVVSATDFGIPLSNTTRVLVRVGISQNPPIPRGHQNMENL 706
Qy      | 503 PIFVSTPFOATVLESVPLGLVLAHQALDADADNARLERLAGVHDPPFTTNGTGM 562
Db      | 707 PIIIGPSPYSRISIPENNAAGSNVTSASHDPD-GIDSLRRIIVGANDNEIDEISGLITV 765
Qy      | 563 SVAAELDRE-EVDPYFSGVEARDHGT-ALTASASVTVLVNDNPPFTOPEYTVRLN 620
Db      | 766 SPQARIDROSNMNSFEIIVAVDSCTPIPETATTYYVAVKDIINDERPKRQNSYATYVS 825
Qy      | 621 EDAAVGTSVTVTVSAVDRDAHSVITTYQITSGNTRNRPSTISQSGGLVSLALPLDYKLERO 680
Db      | 826 ERTAVGESVLRVAKIDKLSKLEYSMV---GPVAATTKAG---VSJANRSNYRLOEA 877
Qy      | 681 YVLAVTASD---GT-RQDTAQV---VNTDANTH-----RP 710
Db      | 878 FRVDSQSGELFVNGTILHVDPAIILFTVGRDINARVDEQVDTTEVTYVQSFQDTNP 937
Qy      | 711 VFQSSHYT-----VNVNEDRPAGTTVVLISATDEDTGENARITYF---MEDSIPQFRI 760
Db      | 938 VFRNTGWTSSRPVIDYKIKEMPIDALPILQABDPTRO--PITSFELIEPKQVDYFOY 995
Qy      | 761 DADTGAVTTOAELDYR---DOVSYTLATITARDNGIPQKSDTYLLEILVNDVNDNAPQL 816
Db      | 996 AERTGEVILKRLDYALGDTGPEFELQVRA-NSADRRQSTSVSRVNIIVENVNDNSRFE 1054
Qy      | 817 RDSYQGSVVEDVPEPFTSVLOISATDRDGLNGR-----VPTYQGDGDDGDFIVBS 868
Db      | 1055 RNSYQATIIENRPHRVRIRVRALDKDAVLNARDERLGYHKIITYSLOG--EHAMLFDINN 1112
Qy      | 869 TSG--IVRTLRLDRENVAVYLRAYAVDK-GMP--PARTMEVTVTVLVNDNPPVFRQ 923
Db      | 1113 TTGEIIVASGQITDRERTRIQIQIKABSPGRFTAKOSVVELQIEVLDDVNDNABEFTQ 1172
Qy      | 924 DEFDFVEENSPIGLAVARVATADPDEGTAQIMYOIV-EGNIPVFOLDIFSGELTALV 982
Db      | 1173 KKSSTVIPENAOIDSFVLQLEAVDADGEGEVRLEYLVNVEGEANGLFKIDTKSGLISTR 1232
Qy      | 983 DLUDYEDRPE-YTVIYQ-----TSAPVSRAT---VHVRLLDRNDNPPV-----LGNF 1026
Db      | 1233 NLTKGRAPHYVILVRAQDNGNMPKOPTLSTDTVIRIYIGVSDANDGVYFLSPRVQOM 1292
Qy      | 1027 EILFNMYVTNRSSSPFGAIGRVPADPD---ISDLYTSFERGN-ELSLVILNASTGB 1081

```

```

Db      | 1293 ANTENAVT-----GAPVFOVIASDPDDDESTPSTITTYRILPDTPPDAEAFIDAHSGI 1345
Qy      | 1082 LKLSRALDNNRPLEAIMSVLVSDBGVHSAVTAQCALRVITITDEMITSITLREDMSP--- 1138
Db      | 1346 ITRQSMO--RETQWYRILLEVSDNGPKQKQSATRILQIA-----ILDVDHPRFA 1395
Qy      | 1139 -REFLSPL-LGFLQAVATLA-----TPPHVVV-----FNVQ----- 1171
Db      | 1396 READAGPLSMVSAREEPACTIVGNSPALDEDLGENAALIDYVITIDGNNBQLFTIERNNESI 1455
Qy      | 1172 -----DTAPGHIINVS-LSVQPPGPGGPPFLPSEDLQERLYINRSLTAISA 1221
Db      | 1456 AILKTKKPIDRQVBSFTLITKCLKLGBGYTFIDPT-----DRQDLSHRLNV----- 1505
Qy      | 1222 QRVLPEDN-----ICLRPCENYMRCVSVLRFSSAPFIAS--SVLFRP 1265
Db      | 1506 -RVLDDIDNLPKREQDPVTGIRINVPIDTVVTTLKASDADAAPVGLSIEVTFVP 1562

```

Search completed: February 11, 2004, 15:57:08  
Job time : 96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 13, 2004, 19:27:01 ; Search time 11477 Seconds  
(without alignments)  
6189.944 Million cell updates/sec

Title: US-09-916-849a-3  
Perfect score: 15545  
Sequence: 1 MESPATGVLPPTPPPLLL.....AGTVDSSGSEFLPFNFILH 2923

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=csn2.1/USPTO.spool/US09916849/rnat.11022004.143816.20951/adb.query.fasta.1.3079  
-DB=EST -QFMT=fastap -SUPPLY=est -MIMATCH=0.1 -LOOPT=0 -LOOPTXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NOR=ext -HEADSIDE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09916849 @CN 1 1 9354 @rnat.11022004.143816.20951 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gse\_hum:\*  
18: em\_gse\_inv:\*  
19: em\_gse\_pin:\*  
20: em\_gse\_vrt:\*  
21: em\_gse\_fun:\*  
22: em\_gse\_mam:\*  
23: em\_gse\_mus:\*  
24: em\_gse\_pro:\*  
25: em\_gse\_rtd:\*  
26: em\_gse\_pug:\*  
27: em\_gse\_vrl:\*  
28: gb\_gse1:\*

29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1598.5	10.3	1076	12	BM480097
2	1578	10.2	896	13	BU173510
3	1523	9.8	907	13	EX409742
4	1477	9.5	990	13	BK460356
5	1464.5	9.4	846	14	CA751074
6	1454.5	9.4	816	13	BU701387
7	1451	9.3	1039	29	CNS0385Y
8	1387	8.9	877	13	EX409741
9	1369	8.8	839	14	CB520719
10	1346	8.7	908	13	BU857213
11	1335.5	8.6	928	12	BU655671
12	1329	8.5	789	14	CB249583
13	1300	8.4	767	14	CB246025
14	1299.5	8.4	893	13	BU849091
15	1284	8.3	812	12	BG915872
16	1281	8.2	694	14	CD350873
17	1267.5	8.2	763	14	CD355642
18	1254	8.1	890	13	BQ920714
19	1240	8.0	745	14	CA328479
20	1180	7.6	598	9	AL704278
21	1180	7.6	719	14	CB245128
22	1173	7.5	597	14	CA891550
23	1173	7.5	723	14	CD350951
24	1170	7.5	738	14	CA510833
25	1168	7.5	833	29	CNS038ZT
26	1166	7.5	722	12	B1456543
27	1141	7.3	875	12	B1693814
28	1121.5	7.2	695	9	AU116945
29	1118	7.2	767	13	BU705756
30	1115	7.2	980	29	CNS04381
31	1114	7.2	662	9	AL042580
32	1101	7.1	861	29	CNS0320M
33	1094	7.0	716	14	CD351314
34	1088.5	7.0	927	13	BQ679970
35	1081.5	7.0	906	13	BQ677400
36	1078	6.9	659	14	BY727867
37	1074	6.9	647	13	BU703750
38	1072.5	6.9	930	13	BQ676596
39	1063	6.8	640	9	AM370778
40	1038	6.7	714	12	B1078247
41	1035	6.7	4075	9	AK048080
42	1011	6.5	616	9	AW748496
43	1007	6.5	660	10	BE226430
44	1003	6.5	907	14	CA471519
45	997	6.4	1137	10	BF121073

#### ALIGNMENTS

RESULT 1  
LOCUS BM480097 1076 bp mRNA EST 05-FEB-2002  
DEFINITION AGENCOURT 6468579 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5574047  
5' mRNA sequence.  
ACCESSION BM480097  
VERSION BM480097.1 GI:18529139  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1076)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LML at:  
<http://image.llnl.gov>  
 Plate: LLM1322 row: a column: 24  
 High quality sequence start: 29  
 High quality sequence stop: 666.  
 Location/Qualifiers

FEATURES  
 source

1. 1076  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5574047"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_68"  
 /note="Organ: small intestine; Vector: pCMV-SPORTS;  
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;  
 oligo-dT primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH MGC library."  
 BASE COUNT 215 a 357 c 310 g 193 t I others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.02e-92 Length: 1076  
 Score: 1598.50 Matches: 319  
 Percent Similarity: 91.71% Conservative: 13  
 Best Local Similarity: 88.12% Mismatches: 18  
 Query Match: 10.28% Indels: 13  
 Gaps: 4

US-09-916-849a-3 (1-2923) x BM480097 (1-1076)

QY 1930 GlycerineSerArgValCysAspProGluAspGly-GlnCysAspProGlySerProGlyVal 1949  
 DB 1 GGTGGATTTCGGGAATCTGACCCCGCTCGCTCCAGTCCATGCAAGCCAGGTGT 60  
 QY 1949 111eGlyArgGlnCysAspArgCysAspAsnProPheAlaGluValThrThrAsnGlyC 1969  
 DB 61 CATGGGGGTGAGGTGACCCCTGTGACACCTTTTGTGAGTCAACCAATGCTGTG 120  
 QY 1969 eGluValAsnTyrAspSerCysProArgAla11eGluAlaGly11eTTPTPProArgTh 1989  
 DB 121 TGAAGTGAATTATATACAGCTCCCAACGAGCATTTGAGGTGGATCTGTGGCCCCGTAC 180  
 QY 1989 rArgPheGlyLeuProAlaAlaProCysProGlyGlySerPheGlyThrAlaValAr 2009  
 DB 181 CCGCTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 QY 2009 gHicCysAspGluHisArgGlyTTPLeuProProAsnLeuPheAsnCysThrSer11eTh 2029  
 DB 241 CCACGTGATGAGCAGCGGGGTGGCTCCCAAACTCTTCAATGCAAGTCACTAC 300  
 QY 2029 rPheSerGluLeuGlyGlyPheAlaGluArgLeuGlnArgAsnGluSerGlyLeuAsp 2049  
 DB 301 CTTCTCAAACTGAAGGGCTTCTGCTGAGCGGCTACAGCGGAATGATGATGAGCTTACGCTC 360  
 QY 2049 rGlyArgSerGlnGlnLeuAlaLeuLeuLeuArgAsnAlaThrGlnHisThrAlaGlyT 2069  
 DB 361 AGGGGCTCCCAAGAGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 QY 2069 rPheGlySerAspValValAlaTyrGlnLeuAlaThrArgLeuLeuAlaHisGluSe 2089  
 DB 421 CTTGGGAGCAGCAGTCAAGTGTGGCTTACAGCTGGCCAGCGGGCTGCTGGCCAGAGAG 480

QY 2089 rThrGlnArgGlyPheGlyLeuSerAlaThrGlnAspValHisPheThrGluAsnLeu 2109  
 DB 481 CACCCAGCCGGGCTTTGGGCTGTCTCCACACAGACGTCGACTTCACTGAGAAATGCTGT 540  
 QY 2109 uArgValGlySerAlaLeuLeuAspThrAlaAsnTyrAsnGlySerGlyLeu11eGln 2129  
 DB 541 GCGGGTGGGACAGGCGCTCTGAGACACAGCCAAAGCGGACCTGGAGCTGATTCAGCA 600  
 QY 2129 nThrGlnGlyGlyThrAlaThrLeuLeuGlnHisSerGlyAlaTyrAlaSerAlaLeu 2149  
 DB 601 GACAGAGGGTGGACCGCTGCTGCTGCAAGCATATAGAGCTTACGCAAGTCCCTGG 660  
 QY 2149 aglAsnMetArgHisThrThrLeuSerProPheThr11eValThrProAsn11eVal11 2169  
 DB 661 CCAAGAACATGCGGACACCTTACCTTAAGCCCTTCAACATGTCAGGCCCACTTCTCAT 720  
 QY 2169 eSerValValArgLeuAspGlyAsnPheAlaGlyAlaGlyLeuProArgTyrGluAl 2189  
 DB 721 CTCCTGATGCGCTTGTGACAAAGAACTTGTGCGGCGCAAGCTCCCGCTACAGAGC 780  
 QY 2189 AlauArgGlyGluGlnProProAspLeuGlnThrThrVal11eLeuProGluSerValPh 2209  
 DB 781 CTTGCTGGGAGACACCCCGGACCTTACAGACAGTCATTTCTGAGTCTGTCTT 840  
 QY 2209 eArgGlnThrProProValValArgProAlaGlyProGlyGlu-AlaGlnGluProGlu 2229  
 DB 841 TCAGAGAGAGCGCCCGCTGTGAGCCGCAAGCCCGGAGAAAGCCAGGAGCCAGAGCCAG 900  
 QY 2229 lu-LeuAlaArgArgGlnArgArg-HisProGluLeuSer---GlnGlyGluAlaVal 2247  
 DB 901 AACCTGCGCCGTGAAAGGACGAGAACCCGAGTTTACCAACCGGGGAGACGTGTGGC 960  
 QY 2247 aSerVal11e11e-----TyrArgThrLeuAlaGlyLeuLeuProHisAsnTyr 2263  
 DB 961 AACGATTCATTCAACCCCAACCCCTGGGGGGGTATTGCTCTTAACT-----TA 1014  
 QY 2263 rAspProAspGlyArgSerLeu---ArgValProGlyArgPro11e11eAsnThrPro 2281  
 DB 1015 TGACCTGAAAGAACCC-CACATTTGAAAGATGCCCAAGCCCGATCATCATCACCC 1071

RESULT 2  
 BUI73510 896 bp mRNA linear EST 04-SEP-2002  
 LOCUS BUI73510 7939348 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:6172948  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BUI73510.1 GI:22687494  
 VERSION BUI73510.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 896)  
 TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LML at:  
<http://image.llnl.gov>  
 Plate: LLM13543 row: 1 column: 05  
 High quality sequence stop: 677.  
 Location/Qualifiers

FEATURES  
 source  
 1. 896  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

```

/clone="IMAGE:6172948"
/issue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH MGC 67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."
BASE COUNT      169 a      282 c      267 g      177 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      1,66e-91      Length:      896
Score:          1578.00      Matches:      295
Percent Similarity: 98.99%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:      10.15%      Indels:      2
DB:              13      Gaps:      0

US-09-916-849a-3 (1-2923) x BU173510 (1-896)

QY      1351 PhelysCyAspCyAspSerSerGlyAspPheGluValProTyrCyAspGlnValThrTharXy      1370
DB      1 TTCAAGTGGATGATGCTGCACTTGGAGACTTGGAGAGCCCTTACTGCGAGGTGACCAAGCCG      60

QY      1371 SerPheProAlaHisSerPheIleThrPheArgGlyLeuArgGlnArgPheHisPheThr      1390
DB      61 AGCTTCCCGCCGCACTCTTCACTACCTTCCGGGCTGGCCAGGCTTCCACTTCACTTCACTC      120

QY      1391 LeuAlaLeuSerPheAlaThrIlysgGluArgArgGlyLeuLeuLeuTyrArgGlnArgPhe      1410
DB      121 CTGGCCCTCTCTTGGCCCAAGAGGCGCGGCTTCTGTGTACATGAGGGCGTTC      180

QY      1411 AengLulYrHISAspPheValAlaLeuGluValIleGlnGluGlnValGlnLeuThrPhe      1430
DB      181 AATGAGAGCATGACTTGTGGCTTGGCCCTCGAGGTGATTCAGAGAGAGGTCCAGCTCACCTTC      240

QY      1431 SerAlaGlyGluSerThrThrThrValSerProPheValProGlyValSerAspGly      1450
DB      241 TCTGCAAGGGAGTCAACACCAACGAGGTGCTCCCATTCGTCCGAGAGAGTCAAGTATGAGC      300

QY      1451 GlnThrPheIleThrValGlnLeuLeuYrTyrTrpAnlyProLeuLeuGlnGlyGlnThrGlyLeu      1470
DB      301 CAGTGCATACAGGTGAGCTGAATACATCAATTAAGCCACTGTGGGTCAACAGGCTC      360

QY      1471 ProGlnGlyProSerGlnGlnValAlaValAlaValThrValAspGlyCyAspThrGly      1490
DB      361 CCAAGAGGCGCCATCAGAGCAAGAGGTGCTGTGATCCCTGATGCTGTATACACAGGA      420

QY      1491 ValAlaLeuArgPheGlySerValLeuGlyAsnTyrSerCyAlaAlaGlnGlyThrGln      1510
DB      421 GTGGCTTGGCTGGATCTGTCTGGGCACTACTCTGTGTGGCCAGGAGCCAGC      480

QY      1511 GlyGlySerIlySerIleuAspLeuThrGlyProLeuLeuGlnGlyValProAsp      1530
DB      481 GGTGGCAGCAAGAAAGTCTGTGATTCAGAGGGGCCCTCTACTAGCGGGGTGCTGAC      540

QY      1531 LeuProGluSerPheProValArgMetArgGlnPheValGlyCyMetArgAnleuGln      1550
DB      541 CTGGCCGAGAGCTTCCCACTCCGAATGCGGCAAGTGTGTGGCTGTCAATGGGAACTTGCAG      600

QY      1551 ValAspSerArgHisIleAspMetAlaAspPheIleAlaAsnAnGlyThrValProGly      1570
DB      601 GTGGACAGCGGCACTAGACATGGGTGACTTCACTTGCACAAATGGCACTGGCTGGC      660

QY      1571 CySPeRoAlaIlyLeuAsnValCyAspSerAspThrCyHisAsnGlyIlyThrCyAspVal      1590
DB      661 TCCCTGTGCAAGAAAGACGTGTGTGACACCAACTTGCACAAATGGGGGCACTTGGCGG      720

QY      1591 AengLulYrAspAlaPheSerCyGluCyProLeuGlyPhe-GlyGlyIlySerCyAla      1610
DB      721 AACGATGGGAGCGCTTCAAGCTGCCAGTCCCTCTGNGCTTGGGGGGCAAGAGCTGCC      780

```

```

QY      1610 aGlnGluMetAlaAsnProGlnHisPheLeuGlySerSerLeu-ValAlaThrPheIlyL      1630
DB      781 CCAAGAAATGGCAATTCACAGCACTTCTGGGAGAGCGCTGTGGCTGGATGAGCC      840

QY      1630 eusSerLeuProIleSerGlnProTyrIlyLeuSerLeuMetPheArgThr      1646
DB      841 TCTGCTGCCATTCACATCCCAAGCCCTGTGATCTTCAAGCTCATGTTCGACAG      890

RESULT 3
BX409742
LOCUS      BX409742      907 bp      mRNA      linear      EST 13-MAY-2003
DEFINITION      BX409742 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION      CS0DF013YE18 5-PRIME, mRNA sequence.
VERSION      BX409742
KEYWORDS      BX409742.1 GI:30653001
SOURCE      EST.
ORGANISM      Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 907)
AUTHORS      Li, W.B., Gruber, C., Jessup, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 64.f For more
information about this cluster, see http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF012ZG07_Af01112_2&cluster=64.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradey Avenue Genoscope sequence ID : CS0BAF012ZG07_Af01112_2.
FEATURES
source
location/Qualifiers
1..907
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF013YE18"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo (dt) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and KcoRI sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT      148 a      322 c      238 g      198 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      5.8e-88      Length:      907
Score:          1523.00      Matches:      297
Percent Similarity: 98.67%      Conservative: 0
Best Local Similarity: 98.67%      Mismatches: 2
Query Match:      9.80%      Indels:      4
DB:              13      Gaps:      0

US-09-916-849a-3 (1-2923) x BX409742 (1-907)

QY      2433 AlaAspLeuProPheAlaCyThrValIleAlaIleuLeuHisPheLeuTyrLeuCy      2452
DB      7 GCTACCTCCCTTTTGGCTGACAGTCACTTCCATCTCGCTGCACTTCTTACTCTGCG      66

QY      2453 ThrPheSerTPAlaLeuLeuGlnAlaLeuHisLeuTyrArgAlaLeuThrGlnValArg      2472
DB      67 ACTTTTCTGTGGCTGTGCTGAGAGGCTTGACCTGTACCGGGCACTCACTGAGGTGGC      126

QY      2473 AspValaLeuThrGlyProMetArgPheTyrTyrMetLeuGlnGlyTyrGlyValProAlaPhe      2492
DB      127 GATGTCAACACCGGCCCATGCGCTTCTACTACATGCTGGGCTGGGGGCTGCTGCTTC      186

```

```

Qy 2493 IleThrGlyLeuAlaValGlyLeuAspProGluGlyTyrGlyAsnProAspPheCysTyr 2512
    |||
Db 187 ATCAAGAGGCTACCGGCGGCTGAGACCCGAGGAGCTACCGGAACTGACTTCTGCTG 246
Qy 2513 LeuSerIleTyrAspThrLeuIleTyrSerPheAlaGlyProValAlaPheAlaValSer 2532
    |||
Db 247 CTCCTCCATCTATGACACGCTCATCTGGAGTTTGCTGGCCCGGCTGCTTGGCGCTG 306
Qy 2533 MetSerValPheLeuTyrIleLeuAlaAlaArgAlaSerCysAlaAlaGlnGlnGly 2552
    |||
Db 307 ATGAGTGTCTTCTGACATCTGCGCGCCCGGCTCTCTGCTGCTGCGCAGCGGAGGC 366
Qy 2553 PheGluValSerGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeu 2572
    |||
Db 367 TTGAGAGAAAGAGTCTCTGCTGCGGCTGAGCCCTCTTCCGCTCTCTGCTGCTG 426
Qy 2573 SerAlaThrTyrLeuLeuAlaLeuLeuSerValAsnSerAspThrLeuLeuPheIleTyr 2592
    |||
Db 427 AGCGCCACGTGGCTGCTGCGACCTGCTCTGTCAACGAGACACCTCTCTTCCACTAC 486
Qy 2593 LeuPheAlaThrCysAsnGlyIleGlnGlyProPheIlePheLeuSerTyrValValLeu 2612
    |||
Db 487 CTCCTGCTACCTGCAATGATGATCAGGCGCCCTTCACTCTCTCTCAATGATGCTT 546
Qy 2613 SerIleGlyValArgIleAlaLeuValLeuAlaCysSerArgIleProSerProAspPro 2632
    |||
Db 547 AGCAAGAGAGTCCGGAAGACATCACTGAGCTGCGCGCCGCAAGCCGACCTGACCT 606
Qy 2633 AlaLeuThrThrIleYsrSerThrIleThrSerSerTyrAsnGlyProSerProTyrAlaAsp 2652
    |||
Db 607 GCTCTGACCAACCAAGTCCACCTGACCTGCTCTCAACAGTCCGCGCCCTCAAGCAGAT 666
Qy 2653 GlyArgLeuTyrGlnProTyrGlyAspSerAlaGlySerLeuIleSerThrSerArgSer 2672
    |||
Db 667 GGGGGGTGTACAGCCCTACGAGACTGCGCGGCTCTTGTGACACAGACCAATGCTG 766
Qy 2673 GlyIleSerGlnProSerTyrIleProPheLeuLeuArgGluGlnSerAlaLeuAspPro 2692
    |||
Db 727 GGCAGAGTCAAGCCGACCTACATCCCTTCTGCTGAGGAGAGTCCGACGACCACT 786
Qy 2693 GlyIleGlnIleProProGlyLeuGlyAspProGlySerLeuPheLeuGlnGlyIleAspGln 2712
    |||
Db 787 GGGCAAGGGGCCCCCTGCGCTGGGGATCCAGGCGCTGCTCTGCGGAGGTCAAGCAG- 845
Qy 2713 GlnHisAspProAspThrAspSerAspSerAspLeuSerLeuGlnAspAspGlnInsrgly 2732
    |||
Db 846 CAGCATATCTCTGACAGCACTGACGACGACGACGACGACGACGACGACGACGACGACG 904
Qy 2733 Ser 2733
    |||
Db 905 TCT 907

```

RESULT 4  
BX460356  
LOCUS  
DEFINITION  
ACCSSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX460356 990 bp mRNA linear EST 22-MAY-2003  
BX460356 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
CSODP013YE18 5-PRIME, mRNA sequence.  
BX460356  
BX460356.1 GI:31035192  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 990)  
Ll,W.B., Gruber,C., Jeejee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 64.f For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODP013BC09QPl&cluster=64.f>. Contact : Feng Liang Email : [liang@lifetech.com](mailto:liang@lifetech.com) URL : <http://fulllength.invitrogen.com/invitrogen> Paraday Avenue Genoscope sequence ID : CSODP013BC09QPl.

FEATURES  
SOURCE  
1. 990  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODP013YE18"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT\_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 193 a 318 c 273 g 195 t 11 others  
ORIGIN

Alignment Scores:  
Pred. No.: 5,89e-85 Length: 990  
Score: 1477.00 Matches: 302  
Percent Similarity: 97.73% Conservative: 0  
Best Local Similarity: 97.73% Mismatches: 6  
Query Match: 9.50% Indels: 4  
DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BX460356 (1-990)

```

Qy 2159 ProPheThrIleValIleThrProAsnIleValIleSerValValArgLeuAspIleGlyAsn 2178
    |||
Db 64 CCTTCACACCAACGTCACCGCCCAATTCATCTCCGTAGTCCGCTTGACAAAGGAGAAC 123
Qy 2179 PheAlaGlyValAlaLeuProArgTyrGlnAlaLeuArgGlyValGlnIleProProAspLeu 2198
    |||
Db 124 TTGCTGGGGCCAAAGTGGCCCGGCTACAGAGCCCTGCGTGGGAGAGCCGCCGACCTT 183
Qy 2199 GlnThrThrValIleLeuProGluSerValPheArgIleThrProProValValArgPro 2218
    |||
Db 184 GAGACAAACGATCTCTGCTGATGCTGCTTCAAGAGACCCGCCGCTGATCAGGCC 243
Qy 2219 AlaGlyProGlyValAlaGlnIleProGluGlnIleValAlaArgGlnArgIleHisPro 2238
    |||
Db 244 GCGAGGCCCGGAGAGGCCCGGAGGCGAGGAGGCTGGCAGCGGACAGCGGACGCCG 303
Qy 2239 GluLeuSerGlnGlyValAlaValAlaSerValIleIleTyrArgThrLeuAlaGlyLeu 2258
    |||
Db 304 GACCTGAGCGGAGGAGGCTGTGGCCAGCGCTGATCTACCGACCTTGGCCGGGCTA 363
Qy 2259 LeuProHisAsnTyrAspProAspIleValArgSerLeuArgValProIleIle 2278
    |||
Db 364 CTGCTCATTAATGACCTGACAGCGCAGATKA-NGAGTCCCAAGCGCCGATCATC 422
Qy 2279 AsnThrProValValSerIleSerValHisAspAspGlnGlnLeuLeuProArgAlaLeu 2298
    |||
Db 423 AACACACCCGTGTAGCATCAGCGCTCCATGATGATGATGATGATGATGATGATGATG 482
Qy 2299 AspIleProValThrValGlnPheArgLeuLeuGlnIleGlnIleGlnIleGlnIle 2318
    |||
Db 483 GACAAACCGGTACAGGTGAGTTCGGCTGCTGAGAGACAGAGAGGAGGACCAAGCCATC 542
Qy 2319 CysValPheTyrAsnHisSerIleLeuValSerGlyThrGlyIleTyrSerAlaArgGly 2338
    |||
Db 543 TGTGCTTCTGGAACCATTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
Qy 2339 CysGluValValPheArgAsnGlnIleValSerCysGlnCysAsnHisMetThrSer 2358
    |||
Db 603 TGTGAAGTCTCTTCCGCAATAGACCGTCAAGTCTGCAATGACCAATGACGAC 662

```

QY 2359 PheAlaValIleuMetaspValSerArgArgIuAenGlyGluIle-LeuProIeuIyStH 2378  
 DB 663 TTCGCTGTCTCATGACGCTTCTCGCGGAGAAAGGAGATCCTTCCACTGAAAGAC 722  
 QY 2378 rleuThyTyValAlaLeuGlyValThrlleuAlaIleuLeuLeuThPhphephe 2398  
 DB 723 ACTGACATACGTGGCTTCTGAGTGTACCTTGGCTTCTGCTCACTTCTTCTCT 782  
 QY 2398 uThrlleuLeuArgIleLeuArgSerAngInHsgIylIeArgArgAsnLeuThAlaI 2418  
 DB 783 CACTCTCTTCCGATCTCGGCTCCAAACCAACGCGCATCCGATCCTGACAGCTTC 842  
 QY 2418 AleuGlyLeuAlaGlnLeuValPheLeuLeuGlyIleAngInAlaAspLeuProPheAl 2438  
 DB 843 CCGGGCCCTGGCTGACCTGCTTCTTC-CTGGAAATCAACAGGCTGACCTCCCTTTC 901  
 QY 2438 acyThValIleAlaIleLeuLeuHIsPheLeuTyrlauCyThPhSerTrAlaIe 2458  
 DB 902 CTGCAACATCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 QY 2458 uLeuGlnAlaLeuHIsLeuTyrlArg 2466  
 DB 961 GCTGGAGGCTTGCACATGTACCG 985  
 RESULT 5  
 CA751074 846 bp mRNA linear EST 27-NOV-2002  
 LOCUS UI-M-FOO-cdm-p-08-0-UI.r1 NIH BMAP\_FOO Mus musculus cDNA clone  
 DEFINITION IMAGE: 6830553 5', mRNA sequence.  
 CA751074  
 VERSION CA751074.1 GI:25577983  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 846)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsrbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNI at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pyx-5.  
 FEATURES  
 source  
 1.846  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE: 6830553"  
 /rname\_type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /note="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I;  
 Site 2: Not I; The library was constructed according  
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pyx-Asc vector. The library tag

sequence located between the Not I site and the polyA tail  
 is TCAGAGACC. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemlin Chlin, Ph.D.,  
 program coordinator."

BASE COUNT 172 a 240 c 247 g 184 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,056-84 Length: 846  
 Score: 1464.50 Matches: 258  
 Percent Similarity: 92.61% Conservative: 5  
 Best Local Similarity: 90.85% Mismatches: 19  
 Query Match: 9,42\$ Indels: 5  
 DB: 14 Gaps: 1

US-09-916-849a-3 (1-2923) x CA751074 (1-846)

QY 1731 GTPAATGLeuHIsGlyLeuHIsLeu-SerAsnIleThrValGlyIleProGlyPr 1750  
 DB 3 GGCCCTCGGCTGACATGACATGACATGACATGACATGACATGACATGACATGAC 62  
 QY 1750 cAlaGlyIleValAlaArgGlyPheArgGlyCylLeuGlnGlyValArgValSerAspTh 1770  
 DB 63 AGCCAGCGAGAGCCCGTGGCTTCCGGGCTGTTGACGGGTGCAGGGTAAGCGAAGC 122  
 QY 1770 rProGlnGlyValAAsnSerLeuAspProSerHIsGlyGlnSerIleAAsnValGlnGln 1790  
 DB 123 ACCTAGAGGTATACACAGATCTAGATCCAGCCGCGGAGAGACATCATGTGACACAG 182  
 QY 1790 yCysSerLeuProAspProCyAspSerAsnProCyAspProAlaAAsnSerTyCySerAs 1810  
 DB 183 CTGTAGCTGGCCAGATCCCTGTGACTCGAATCCATGACCTTACCAACACTACTGACGCA 242  
 QY 1810 nAspTrpAspSerTySerCySerCyAspProGlyTyTrpGlyAspAsnCySerThAs 1830  
 DB 243 TGACTGGAGACATATCTTGTAGCTGTGTCTTGTAGTATCATGTGTGACACTGTACAA 302  
 QY 1830 nValCyAspAspLeuAsnProCyAspGlnHIsGlnSerValCySerThArgGlyProSerAlaPr 1850  
 DB 303 TGTGTGATCACTGAAACCATGAGACACAGTCCGTGTGATCCGAAACCAATACACC 362  
 QY 1850 cHIsGlyTyTrpCyGlnGlyCyAspProAsnTyrlLeuGlyProTyCyGlnTyTrpArgTl 1870  
 DB 363 CCACGGCTACATCTGCGAGTGTTCACAAATTAACCTTGGCCATATTGTGACACAGAT 422  
 QY 1870 cAspGlnProCyAspProArgGlyTyTrpGlnHIsProThrCyGlnTyProCyAsnCyAs 1890  
 DB 423 TGACCAACCTTGGCCCCCGTGGCTGGAGGACACCCACATGTGTCCATGTCAACTGTGA 482  
 QY 1890 rValSerTyGlyPheAspProAspCyAspAsnTyrlSerGlyGlyGlyGlyGlyGlyGly 1910  
 DB 483 TGTTCAGCAAAAGCTTTGACCCAGATTCGACAAAGCAAGTGGGAGGTGCATGCAAGA 542  
 QY 1910 uAsnHIsTyrlArgProProGlySerProThrCySerLeuLeuCyAspCyTyrlProThrGl 1930  
 DB 543 GAATCACTACCGGCCCCCGGAGCCCACTGTGTCTGTGTGTGACTGTACCCCACTGG 602  
 QY 1930 ySerLeuSerArgValCyAspAspProGlnAspGlyGlnGlyProCyGlyAspProGlyValIl 1950  
 DB 603 TTTCTTGTCCAGCTGTGACCCCGAGAGACGCGCAAGTCCGTCAAGACCTGGAGTAT 662  
 QY 1950 eGlyArgGlnCyAspAspArgCyAspAspProPheAlaGlnValThrThAsnGlyCyGln 1970  
 DB 663 TGGGGGTACAGTGTACCGCTGTGAC-CCTTTGCTAGGTCAACCAATAGCTGTGA 721  
 QY 1970 uValenTyrlAspSerCyProArgAlaIleGlnAlaGlyIleTyTrpProArgThAr 1990  
 DB 722 AGTGAATTAACACACACTCCCAACGCGCATAGAGCTGTGGATCTGGTGGCCGC-ACCG 780  
 QY 1990 gPheGlyLeuProAlaAlaIleAlaProCyProGlySerPheGlyThrAlaValArgHIs 2010



Db 781 GTTCGGGCTA--CTGCTGCTGCCCTGCCCANAGCTCTTGAGACTGTGTGCG-CA 836  
 QY 2010 sCyaaPglu 2013  
 Db 837 CTGTGATGAG 846

RESULT 6  
 BU701387 816 bp mRNA linear EST 09-OCT-2002  
 LOCUS BU701387  
 DEFINITION UI-M-EXO-bvK-1-14-0-UI.r1 NIH\_BMAP\_EXO Mus musculus cDNA clone  
 IMAGE: 5719381 5', mRNA sequence.

ACCESSION BU701387 GI:23625135  
 VERSION BU701387.1  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 816)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

FEATURES  
 source  
 Location/Qualifiers  
 1..816  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE: 5719381"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 15.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_1ib="NIH BMAP EXO"  
 /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then  
 cloned directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GTGCGTGA. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP): 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institutes of Mental Health  
 (NIMH), Hemlin Chn, Ph.D., program coordinator."

BASE COUNT 164 a 256 c 225 g 168 t 3 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,28e-83 Length: 816  
 Score: 1454.50 Matches: 251  
 Percent Similarity: 97.33% Conservative: 4  
 Best Local Similarity: 95.80% Mismatches: 6  
 Query Match: 9.36% Indels: 1  
 Gaps: 1  
 DB: 13

US-09-916-849a-3 (1-2923) x BU701387 (1-816)

QY 1838 GIUHGInServalCyethrArGlyBProSeralAProHISglYrThrCygluCyS 1857  
 Db 3 GAGCAGCAGTCGTGTGTATCCGAAACCCATATACCCGAGCTTACATCTGCAGGT 62  
 QY 1858 ProPdaNTYrLeuGlyProTYrCygluThrAglleabPglProCys---ProArg 1876  
 Db 63 TTACCAAAATTACCTTGGGCAATTGTGAGACAGAGATTGACCAACTTGTGCCCGGANT 122  
 QY 1877 GUYTPTPGLYNhiAProThrCyglYProCysaNCyaaPValSerIyPheabP 1896  
 Db 123 GCGTGTGGGACACCCCAATGTGTCTATGCACTGTATGTACGAGAGGCTTTGAC 182  
 QY 1897 ProaPcyasNlySthSerGlygluCySthiSylSylasNhiSlyrArgProFo 1916  
 Db 183 CCAAGATTGAACACAGACAAAGTGGGAGTGGCCATCCAGAGAAATCACTACGGGCCCCC 242  
 QY 1917 GlySerProThrCySleuLeuCyaaPcySlyrProThrGlySerleuSerArgValCyS 1936  
 Db 243 GGCAGCCCCCACTTGTCTTGTGTACGTGTAACCCCACTGTCTTGTGTCCGAGTCTGT 302  
 QY 1937 AAPProGluAaPGLyGluCyBProCyslySPProGlyValIleGlyArgGluCyaaPArg 1956  
 Db 303 GACCCCGAGAGAGCGCAAGTTCGTGCAAGCTGAGTCAATTGGGCGTCAAGTCAACCC 362  
 QY 1957 CyaaPaSPProPheAlleGluValThrThrAenglyCygluValAsnTyraSPserCyS 1976  
 Db 363 TGTGACAAACCTTTGTGAGGTCAACCAATGCTGTGAATGAATTACAGACAGCTGC 422  
 QY 1977 ProaGlaIleGluAlleGlyIleThrPProaGlyThrArgPhnglyleuProAlaIa 1996  
 Db 423 CCAAGGCAATAGAGGCTGGAGTCTGGGCCCCGAGGCTTGGGCTTCTCTGT 482  
 QY 1997 AlaProCySPProLyGlySerPheGlyThrAlaValArgHisCyaaPGLyHISArgGly 2016  
 Db 483 GCGCCCTGCCCCAAGAGGCTCTTTGGAGCTGTGTGCGCACTGTGATGACAGACAGGGGT 542  
 QY 2017 TPLeuProProaSPleuPheanCySthSerIleThrPheSerGlyleuLyglYpHe 2036  
 Db 543 TGGCTTCCCCCAAACTCTTCACTGACAGTCACTCACTTCTCAGAGCTAAAGGGCTTC 602  
 QY 2037 AlaGluArgleuGluArgleuGluSerGlyleuAaSPserGlyArgSerGlnleuAla 2056  
 Db 603 GGTGAGCGGCTGCAAGAACCAATCAGGCTGAGACTGAGAGCTCCAGAGGGCTAGGC 662  
 QY 2057 leuLeuLeuArgaAlaIthrGlnHISthralaGlyTYrPheGlySerAaPValIySVal 2076  
 Db 663 CTGCTCTCGCAATGCCACCCAGCACACTCTGCTACTTGGGAGATATCAAGGTG 722  
 QY 2077 AlaTYrGluLeuAlaIthrArgleuLeuAlaHISgluSerThrGlnArgGlyPheGlyLeu 2096  
 Db 723 GCTTACCAAGCTGGCACAAGGCTTGTGCTCATGAGAGTCCACAGGGGGCTTGGGCTA 782  
 QY 2097 SerAla 2098  
 Db 783 TCCGCA 788

RESULT 7  
 CNS0385Y/c 1039 bp DNA linear GSS 01-SEP-2000  
 LOCUS CNS0385Y/c  
 DEFINITION Tetradon nigraviridis genome survey sequence T7 end of clone  
 004013 of library G from Tetradon nigraviridis, genomic survey  
 sequence.

ACCESSION AL232207  
 VERSION AL232207.1 GI:7891342  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigraviridis  
 ORGANISM Tetradon nigraviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Tetradon.

REFERENCE  
 AUTHORS Roest Crolius, H., Jallion, O., Dadiya, C., Bouneau, L., Fisher, C.,



## FEATURES

## SOURCE

Location/Qualifiers

1. .877  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF013YE18"  
 /class\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 140 a 315 c 226 g 194 t 2 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 3.06e-79 Length: 877  
 Score: 1387.00 Matches: 275  
 Percent Similarity: 94.50% Conservative: 0  
 Best Local Similarity: 94.50% Mismatches: 16  
 Query Match: 8.92% Indels: 2  
 DB: 13 Gaps: 0

US-09-916-849a-3 (1-2923) x BX409741 (1-877)

QY 2433 AlaAspLeuProPheAlaCyThrValIleAlaIleLeuLeuHisPheLeuYrLeuCyS 2452  
 DB 6 GCTACCTCTCTTGTGCTGCCA-GTCAATGCCATCTGCTGCATCTTCTGCTGCTGCG 64  
 QY 2453 ThrPheSerTrpAlaLeuLeuGluAlaLeuHisLeuYrArgAlaLeuTrpGluValArg 2472  
 DB 65 ACCTTTCTCTGGGCTGCTGGAGAGCCCTTGACCTGTACCGGGCACTCAGAGTGGCG 124  
 QY 2473 AspValAsnTrpArgPheMetArgPheTrpMetLeuGlyTrpGlyValProAlaPhe 2492  
 DB 125 GATGTCAACACGGCCCGCATGCGCTTCTACATACGTGGCTGGCGCGTGGCTGCTTC 184  
 QY 2493 IleThrGlyLeuAlaValGlyLeuAspProGluGlyYrGlyAsnProAspPheCyS 2512  
 DB 185 ATCAAGGGGCTAGCGGCTGGGCTGACCCCGAGGGCTACGGGAACCTGACTTGTGCTGG 244  
 QY 2513 LeuSerIleYrAspTrpIleTrpSerPheAlaGlyProValAlaPheAlaValSer 2532  
 DB 245 CTCTCCATCTATGACAGCTCATCTGAGGTTCCTGCTGGCCGCGTGGCTTGGCGTCTCG 304  
 QY 2533 MetSerValPheLeuYrIleLeuAlaAlaArgAlaSerCyAlaAlaGlnArgGlnGly 2552  
 DB 305 ATGAGTGTCTTCTGTACATCTGCGGCGCGGCTCTCTGTGCTGCCAGCGGCAAGGCG 364  
 QY 2553 PheGluTrpGlyProValSerGlyLeuGlnProSerPheAlaValLeuLeuLeu 2572  
 DB 365 TTTCAGAGAAAGGTCTCTGCTCGGCGCTGACGCCCTCTTCCGCGTCTCTGCTGCTG 424  
 QY 2573 SerAlaTrpTrpLeuLeuAlaLeuLeuSerValAsnSerAspTrpIleLeuPheHisTrp 2592  
 DB 425 AGCCCAAGTGGCTGGCACTGCTCTGTCTCAACAGCGACACCTCTTCTCACTAC 484  
 QY 2593 LeuPheAlaTrpCyAsnCyAlaGlnGlyProPheIlePheLeuSerYrValValLeu 2612  
 DB 485 CTCTTGTACCTGCAATGATCAGAGGCCCTTCACTTCTCTCTCAATGATGATCTT 544  
 QY 2613 SerIleGluValArgAlaLeuLeuLeuAlaCySerArgGlyProSerProAspPro 2632  
 DB 545 AGCAAGAGGTCGGGAAAGACCTCAGCTTGCCTGACGCCGCAAGCCCAAGCCCTGACCT 604  
 QY 2633 AlaLeuTrpTrpIleYrSerTrpLeuTrpSerSerYrAsnCyProSerProYrAlaAsp 2652  
 DB 605 GCTGTGACCAACCAAGTCCACCTGATCTGCTTCAAACTGCCACGCCCTTACGACAT 664  
 QY 2653 GlyArgLeuYrGlnProYrGlyAspSerAlaGlySerLeuHisSerTrpSerArgSer 2672

## DB

665 GGGCGGCTGTACCAAGCCCTACAGACTCGCGCGCTCTTGCACAGACCACTGCTCG 724

## QY

2673 GlyYrSerGlnProSerTrpIlePheProPheLeuArgGluIleSerAlaLeuAspPro 2692

## DB

725 GGGAAAGTCAAGCCGACGCTACATCCCTTCTGTCTAGAGGAGAGTCCGAACTGAAACT 784

## QY

2693 GlyGlnGlyProProGlyLeuGlyAspProGlySerLeuPheLeuGlnGlyAspGln 2712

## DB

785 GGCACAGG-CGCCCTGGCTGNGGATTCAGCCAGGCTTGTCTTGAAGGCAAGACAG 843

## QY

2713 GlnHisAspProAspTrpAspSerAspAsp 2723

## DB

844 CAGCATGATTTCTGTACACACATTCGACAGTAC 876

## RESULT 9

CB520719 839 bp mRNA linear EST 28-MAR-2003

## LOCUS

CB520719

## DEFINITION

UT-M-GIO-cej-j-12-0-UT.r1 NIH\_BMAP\_G10 Mus musculus cDNA clone

## ACCESSION

CB520719

## VERSION

CB520719.1 GI:29354074

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 839)

## AUTHORS

NIH-MGC <http://img.nci.nih.gov/>.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouse1.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

## FEATURES

Location/Qualifiers

1. .839

## SOURCE

/organism="Mus musculus"

## DB

/mol\_type="mRNA"

## DEFINITION

/strain="C57BL/6"

## ACCESSION

/db\_xref="taxon:10090"

## VERSION

/clone="IMAGE: 6840397"

## KEYWORDS

/class\_type="whole brain"

## SOURCE

/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"

## TITLE

/lab\_host="DH10B (T1 phage resistant)"

## JOURNAL

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lemmon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

## BASE COUNT

133 a 280 c 233 g 190 t 3 others

## ORIGIN

Alignment Scores:

Pred. No.: 4.15e-78 Length: 839  
 Score: 1369.00 Matches: 263



DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	362	CTGCACAGTACTCCCAAGAGATGAGGAGCCCAAGGAGCTTGAGCAAGGCCCTCGGCGAGAGAAC	213									
Qy	2794	Pheg1yThrThrAlaValGluSerSerGlyAenGlyAlaProGluGluAlaGluArgGlu	2813									
Db	422	TTTGGGACCAACAGCAAAAGAGGTGTGTGGCAACGGGAGCCCTCGAGAGACGGCTGCGGAG	481									
Qy	2814	AenGlyAenAlaLeuSerArgGluGlySerLeuGlyProLeuProGlySerSerAlaGln	2833									
Db	482	AATGAGAGTGCCTCTCTCGAGAGGGGTCTCCAGGCCCTTCAGGCTCTTTCGCCAG	541									
Qy	2834	ProH1aValGlyLeuLeuValGlyValGlyCysLeuProThrL1aSerGluLysSerSerLeu	2853									
Db	542	CCTCAACAAGGATCTCTTAAGAGAGAGTGTCTGCCCCACATCAGAGAGAGAGAGCGCTC	601									
Qy	2854	LeuArgLeuProLeuGluGlnGlyThrGlySerSerArgGlySerSerAlaSerGluGly	2873									
Db	602	CTGCGGCTCCCTCGAGAGCATGACAGGGGTCTTCGGAGGCTCTCGCTAGTGTAGGAG	661									
Qy	2874	SerArgGly-GlyProProArgProProProArgGlnSerLeuGlnGlnLeuAs	2893									
Db	662	AGCCGGGAGGGGCCCCCTCCCTCCCGCCACCGCCCGGAGAGCTCCAGAGCACTGATA	721									
Qy	2913	ng1yAlaMetProL1aMetSerL1aValAlaGly-ThrValAspGluAspSerSerG	2913									
Db	722	CGGGGTCATGCCCATGCGCATGAGCATCAAGGCAAGCAAGTGTGATGAGGACTGTGAG	781									
Qy	2913	1y 2913										
Db	782	GG 783										
RESULT 11												
LOCUS	B1655671	928 bp	mRNA	linear	EST 12-SEP-2001							
DEFINITION	60328395881 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5327966 5',											
ACCESSION	B1655671											
VERSION	B1655671.1	GI:15563907										
KEYWORDS	EST.											
SOURCE	Mus musculus (house mouse)											
ORGANISM	Mus musculus											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.											
TITLE	1 (bases 1 to 928)											
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/											
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)											
	Unpublished											
	Contact: Robert Strausberg, Ph.D.											
	Email: cgapbs-r@mail.nih.gov											
	Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth											
	Ph.D.											
	cDNA Library Preparation: Life Technologies, Inc.											
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLT)											
	DNA Sequencing by: Incyte Genomics, Inc.											
	Clone distribution: MGC clone distribution information can be											
	found through the I.M.A.G.E. Consortium/LMLT at:											
	http://image.llnl.gov											
	Plate: L1AM1832 row: d column: 15											
	High quality sequence stop: 769.											
FEATURES	Location/Qualifiers											
source	1..928											
	/organism="Mus musculus"											
	/mol_type="mRNA"											
	/strain="NMRI"											
	/db_xref="taxon:10090"											
	/clone="IMAGE:5327966"											
	/tissue_type="tumor, gross tissue"											
	/dev_stage="5 months"											
	/lab_host="DH10B"											

N1H Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."	
BASE COUNT	168 a 302 c 239 g 219 t
ORIGIN	

**Alignment Scores:**

Pred. No.:	6,74e-76	Length:	9
Score:	1335.50	Matches:	2
Percent Similarity:	85.37%	Conservative:	1
Best Local Similarity:	80.49%	Mismatches:	2
Query Match:	8.59%	Indels:	2
DB:	12	Gaps:	3

US-09-916-849A-3 (1-2923) X B1655671 (1-928)

QY	2422	laqlnleuValPheleuLeuG Y  leanglAlaaspLeuProPhealAcyethVal	2441
Db	2	GCCACAGCTGGCTCTTCTCCCTGGGAGATCAACAGAGCTGACTCCCTTTGGCTGTAGCTC	61
QY	2442	leallleleuLeuHisPheLeuYrLeuCyethrPheSerTrpAlaleuGluAla	2461
Db	62	ATCCCTATCTCGCTGCATTTCTGTACTCTGTGACCTTCTCTGGGCTCTGGTGGAGGCC	121
QY	2462	LeuHisLeuYrArGalaleuthrGluValArgaspValAsnthrGlyProMetArgPhe	2481
Db	122	TTACACCTGTAGCCGCGCTCACAGAGGTCCGCGACCTCAATGCAAGTCCCATGCTTTC	181
QY	2482	TYrTYrMetLeuG YrTPG YValProAlaPhe  ethrGlyLeuAlaValAG YLeuAsp	2501
Db	182	TACTACATGCTGGGCTGGGCGGTCCCTGCTTTCATCAAGGCTCCGCTGGTGGCTTGAT	241
QY	2502	ProGluG YrYrG YAsnProAspPheCyethrPheSerIleYrAspThrLeu  etr	2521
Db	242	CCGAAAGGCTATGGGAACCTGACTTCTGGCGCTCTGTGTATATGATACCTCATCTGG	301
QY	2522	SerPheAlaG YrProValAlaPheAlaValSerMetSerValPheLeuYr  leuAla	2541
Db	302	AGCTTCTGTGACCAAGTGGCTTTCGTTTCATGTAGTGTCTTCCGTACATCTCTGCG	361
QY	2542	AlaArG YAserCyValAlaG YAsnArgG YPheG YLeuYs YSG YProValSerG Y	2561
Db	362	GCCGAGAGCTCTGTGTGCTGCCAAGGAGGCTTCCAGAAAGAAAGGAGGCTGTCTCGGC	421
QY	2562	LeuG YnProSerPheAlaValLeuLeuLeuLeuSerAlaThrTrpLeuLeuAlaLeuLeu	2581
Db	422	CTGGGCTCTCTCTCAAGGCTCTCTGTGCTGTAGTGCAAGTGGCTGTGGCACTGCTC	481
QY	2582	SerValAsnSerAspThrLeuLeuPheHisTYrLeuPheAlaThrCYeAsnCYs  leG Y	2601
Db	482	TCTGTCAACAGTGCACCTGCTCTTCCACTCACTCTTCGCTGCTCAATGTGTGCTAG	541
QY	2602	G YrPhePhe  ePheLeuSerTYrValValLeuSerYsG YValAG YValAlaLeuYs	2621
Db	542	GGCCCTTCAATCTTCTCTCTCTAGGTGTGCTGCAAGAGAGTCCGGAAGCACTGAG	601
QY	2622	LeuAlaCySerArgYsPProSerProAspProAlaLeuThrTrpYsSerThrLeuThr	2641
Db	602	TTGCGCTGTAGCCGGAAGCCAGTCCCGACCTCTTGACCACTAAGTCACTGACCC	661
QY	2642	SerSerTYrAsnCYsProSerProTYrAlaAspG YArgLeuTYrG YnProTYrG YAsp	2661
Db	662	TGCTCTATAATCTCCCAAGCCCTCTAGCGAGACGGAAGGCTGTACAGCCTTAGGAGAT	721
QY	2662	SerAlaG YSerLeuHisSerThrSerArgSerG YysSerG YnProSerTYr  ePro	2681
Db	722	TGCGGTGCTCTGTGCACAGTCCAGCCGATCCCGCAAGATCAAGCAATCAT---CCC	778
QY	2682	PheLeuLeuArgG YnG YSerAlaLeuAsnProG YnG YnG YnG YProProG YLeuG YAsp	2701
Db	779	CTTCTTGTGAGGAAGTCCACCTGAA---CTGGCAAGGTTCCTCCCTCGGCGACAGGGAA	837
QY	2702	ProG YSerLeuPheLeuG YnG YAlaAspG YnG YAspProAspThrAspSerAsp	2721









QY 1894 GlyPheASPProASPcysasnlysthrSerglyGluCysHisCyslysgluasnHis-Ty 1913  
DB 2 GGCTTGACCCAGATTGCAACAGACAGAGGGAGTGCACCTGCAGAGGAGATCACTAG 61  
QY 1913 rargProProGlySerProThrCysleuLeuCysaspCysTyProThrGlySerleuSe 1933  
DB 62 CCGGTCCCGGCGAGCCCACTTGCTCTTGCTGTGTGACTGTACCCCACTGGTCTTTGTC 121  
QY 1933 rargValCysaspProGlyuaspGlyGlnCysProCyslyspProGlyValIleGlyValrGly 1953  
DB 122 CCGAGCTGTGAGCCCGAGAGACCGGCAAGTGTCCGTGCAGACCTGCAGAGTCAITGGGCTCA 181  
QY 1953 nCysaspargCysaspasnProPhealagluValIthrThrAsnGlyCysgluValasnTy 1973  
DB 182 GTGTGACCGCTGTGACAACTTTGTGCTGAGTCAACCAACATGCGTGTGAAGTGAATTA 241  
QY 1973 rAspSerCysProArgAlallegluValaglyIleTPTTPProArgThrArgPheGlyle 1993  
DB 242 CGACACCTGCCCCAGGGCCATAGAGGCTGGATCTGGTGGCCCGCAACGGGTTGGGCT 301  
QY 1993 uProAlaalAalAProCysProlysglySerPheGlyThrAlaValArgHisCysaspGly 2013  
DB 302 ACCTGCTGCTGCCCTGCCCTGCAAGGCTCTTTGGGACTGCTGTGCCCACTGTGATGA 361  
QY 2013 uHisargGlyTyrleuProProasnleuPheasnCysThrSerIleThrPheSergIule 2033  
DB 362 GCACAGGGGTTGGCTCCCCCAAACTTCCTCACTGCACGTCAGTCACTTCAGAGCT 421  
QY 2033 uLysglyPheAlaGluargleuGlnargAsnGlyuaspSerglyIleuaspSerglyIle 2053  
DB 422 AAAGGCTTTCGTGAGGGCTGCAAGAGAAAGATCAGCTGAGTCAAGAGAGCTTCCA 481  
QY 2053 ngInleuAlaleuLeuLeuargAsnAlaThrGlnHisThrAlaGlyTyrPheGlySerAs 2073  
DB 482 GAGGCTAGCCCTGCTCTGCGCATGTGCACCGACACACCTTGCGCTACTTGCGCAGTGA 541  
QY 2073 pValIlysvAlAlaTyrGlnleuAlaThrargleuLeuAlaHisGlySerThrGlnArgGly 2093  
DB 542 TGTCAAGGTGGCTTACCAAGCTGCGCAACGGCTTTGGCTCATGAGAGTGCACCGCGGG 601  
QY 2093 yPheGlyLeuSerAlaThrGlnaspValHisPheThrGluAsnleuLeuargValGlySe 2113  
DB 602 CTTTGGGCTATCCCGACACAGATGTACCTTCACGAGAAATCTGCTGAGGGTGGCAG 661  
QY 2113 rAlaleuLeuaspThrAlaasnlysaArgHisTPTGluLeuIleGlnIthrGluGlyGly 2133  
DB 662 CGC-CTCCTGTGATGC-AGCAACAAGAGGCACTGGAACTGATCCAGCAGACAGAGGGTGG 719  
QY 2133 yThrAlaTTPleuLeuGlnHis-TyrGluAlaTyrAlaSerAlaLeuAlaGlnAsnMetA 2153  
DB 720 CAGGCC-TGGCTGCTCCAGACATAATGAGGCTTACGCGCAGCGC-CTTGCCCAAGATATGC 777  
QY 2153 rHisThrTyrleuSerProPheThrIleVal 2163  
DB 778 GGCA-AACTACTTAC-CCCTTACCATCTGC 807

Search completed: February 14, 2004, 05:36:28  
Job time : 11547 secs